Niveau(x) d'enseignement: 0
Centre d'intérêt: 08 Numéro EDUQ : 8170192

Béland, François.

Une méthode d'analyse maintenant disponible en langage APL : l'analyse nominale hiérarchique de fréquences (ANOMHI) / François Béland. -- Québec : Université Laval, Faculté des sciences de l'éducation ; Montréal : Université de Montréal, Département de sociologie, 1979. 124 p. ; 23 cm. -- (Les Cahiers d'A.S.O.P.E. ; 8)

Bibliographie: p. 97-99.

Cote Dewey: 001.422/B426u Cote LC: HA/30

Descripteur(s): *Analyse de données *Analyse de variables *Statistique informatique)

Identificateur(s) de nom: *Asope (Projet) *ANOMHI (Programme informatique)

Identificateur(s) de sujet: *Analyse nominale hiérarchique de fréquences

Une méthode statistique n'est vraiment disponible en sociologie que si elle peut traduire en langage formel les propositions sociologiques à propos d'une distribution d'événements et permettre des manipulations arithmétiques ou informatiques des algorithmes peu complexes. C'est maintenant le cas de l'analyse log-linéaire qui applique à des données nominales réunies dans un tableau de contingence un équivalent de l'analyse de variance. Le but de ce document est de présenter un programme (ANOMHI: analyse nominale hiérarchique de fréquences) permettant d'aborder de façon compréhensive l'étude des tableaux de contingence avec ses raffinements les plus importants.

Deux étapes. D'abord les variantes disponibles de l'analyse log-linéaire sont présentées au moyen d'un exemple, dans le contexte pratique d'une recherche (coefficients d'association, modèles, normalisation des fréquences estimées, analogue de l'analyse de régression, étude de plusieurs échantillons, quasi-indépendance, indépendance régionale). Par la suite, on présente et on décrit le programme ANOMHI (reproduit en annexe), écrit en langage APL.SVA: entrée des données, sous-routines de calcul des coefficients d'association, estimation des fréquences par un ou des modèles non saturés, normalisation des fréquences, les coefficients de détermination et d'efficacité.

Type de document: s
Financement: Bénéficiaire - Université Laval et Université de Montréal
Organisme(s) source(s) - MEQ - DGES
Programme(s) de subvention - FCAC
Auteur de l'analyse: ar

E D U Q 8170192

LES CAHIERS D'A.S.O.P.E.

Volume VIII

# L'ANALYSE NOMINALE HIÉRARCHIQUE DE FRÉQUENCES (ANOMHI): UNE MÉTHODE D'ANALYSE MAINTENANT DISPONIBLE EN LANGAGE APL.

François Béland

A.S.O.P.E.

ASPIRATIONS SCOLAIRES et
ORIENTATIONS PROFESSIONNELLES DES ÉTUDIANTS

Faculté des Sciences de l'Éducation
Université Laval
Québec

Département de Sociologie
Université de Montréal
Montréal

E D U Q 8170192

LES CAHIERS D'A.S.O.P.E.

VOLUME VIII

# UNE METHODE D'ANALYSE MAINTENANT DISPONIBLE EN LANGAGE APL: L'ANALYSE NOMINALE HIERARCHIQUE DE FREQUENCES (ANOHI)

FRANCOIS BELAND

SEPTEMBRE 1979

## L'EQUIPE ASOPE

Directeurs: Pierre W. BELANGER, Université Laval
Guy ROCHER, Université de Montréal

Chercheurs:

| Avant 1978-79 | En 1979-80 |
|---|---|
| Muriel GARON-AUDY | Paul BELAND |
| Denis FAGNAN | Léon BERNIER |
| François BELAND | Robert BEDARD |
| Suzanne VEILLETTE-PERRON | Louise LAFORCE |
| Louise MALO | Pierre ROBERGE |
| Jean RENAUD | Pierre ROSA |
| Sylvie BEAUCHAMP-ACHIM | |
| Suzanne CHOUINARD | |
| Mireille LEVESQUE | |
| Pierre DROUILLY | |
| Christine CORBEIL | |
| Guy LEGAULT | |
| Alain MASSOT | |

Cette recherche a été réalisée grâce à des subventions du programme F.C.A.C. (Formation de chercheurs et action concertée) et de la Direction générale de l'Enseignement Supérieur du Ministère de l'Education du Québec.

L'auteur tient à remercier Jean-J. Fortier pour ses conseils et critiques judicieux.

## TABLE DES MATIERES


INTRODUCTION ..... 1

I. EXEMPLES D'UTILISATION ..... 3

A. LES COEFFICIENTS D'ASSOCIATION ..... 4

B. LES MODELES ..... 15

C. UNE LECTURE PLUS CONCRETE DU MODELE: LA NORMALISATION DES FREQUENCES ESTIMEES ..... 26

D. UN ANALOGUE DE L'ANALYSE DE REGRESSION ..... 30

E. L'ETUDE DE PLUSIEURS ECHANTILLONS ..... 36

F. LA QUASI-INDEPENDANCE ..... 46

G. L'INDEPENDANCE REGIONALE ..... 52

II. ANOMMI: UN PROGRAMME EN LANGAGE APL.SVA POUR L'ETUDE DES TABLEAUX DE CONTINGENCE MULTIVARIES .. 57

A. L'ENTREE DES DONNEES ..... 58

B. LES SOUS-ROUTINES DE CALCUL DES COEFFICIENTS D'ASSOCIATION ..... 67

C. L'ESTIMATION DES FREQUENCES PAR UN OU DES MODELES NON SATURES ..... 70

D. LA NORMALISATION DES FREQUENCES ..... 82

E. LES COEFFICIENTS DE DETERMINATION ET D'EFFICACITE ..... 91

BIBLIOGRAPHIE ..... 95

APPENDICE 1 - LE PROGRAMME ANOMMI ..... 101

# INTRODUCTION

L'analyse des données nominales a connu depuis le début de cette décennie des développements majeurs. Auparavant, les tableaux de contingence étaient soumis soit à un test de $X^2$, dans le cas bi-varié, soit à la procédure d'étude des interactions de P. Lazarsfeld (1968), soit à l'une des nombreuses mesures d'association disponibles (L.A. Goodman et W.H. Kruskall, 1954; 1959; 1963). Maintenant, par l'intermédiaire de l'analyse log-linéaire, il est possible d'appliquer à des données réunies dans un tableau de contingence un équivalent de l'analyse de variance.

L.A. Goodman (1968, 1970) a participé au développement de l'analyse log-linéaire et l'a introduite en sociologie (1972a, b). D'autres auteurs ont abouti à des résultats identiques par l'approche de la théorie de l'entropie (H.H. Ku et S. Kullback, 1968). Mais le début du développement de l'analyse des tableaux de contingence multidimensionnels date de 1935 avec l'article de M.S. Bartlett. A la fin des années '50 et au début des années'60, A.W. Kimball (1954), S.N. Roy et A. Kastenbaum (1956) et M.W. Birch (1963), pour ne nommer que ces auteurs avaient proposé l'essentiel de la théorie de l'analyse des tableaux de contingence par les modèles log-linéaires ou par la partition du $X^2$. Mais ces propositions se présentaient sous une forme rébarbative, peu attirante pour un utilisateur non-spécialisé et étranger à la statistique.

Une méthode statistique n'est vraiment disponible en sociologie que lorsque son utilisation est simple. Elle doit satisfaire

deux conditions. 1) pouvoir traduire en langage formel les propositions sociologiques à propos d'une distribution d'événements et 2) permettre des manipulations arithmétiques ou informatiques des algorithmes peu complexes et facilement accessibles. J.A. Davis (1974) a présenté aux sociologues un exposé pédagogique sur les méthodes log-linéaires. De même, le programme ECTA a rendu disponible sur ordinateur l'essentiel de l'analyse log-linéaire. Nous présentons ici, sous la forme d'un exemple détaillé les variantes disponibles de l'analyse log-linéaire, en plus de certains ajouts originaux. Par la suite, un programme écrit en langage APL.SVA est présenté et décrit en annexe. La présentation simultanée d'un exemple, de la méthode d'utilisation d'un programme entièrement conversationnel et simple d'usage et du programme lui-même devrait permettre au chercheur d'acquérir une vision synthétique de l'utilisation des modèles log-linéaires. En particulier, l'exemple traité présente une stratégie d'utilisation de l'analyse log-linéaire dans le contexte pratique d'une recherche. Ainsi, au départ, un modèle théorique est posé et après l'étude des coefficients d'association, des modifications y sont introduites, et ce, avant même que le test d'hypothèse soit appliqué. Enfin, le test d'hypothèse est mené à la fois sur le modèle théorique et sur le modèle modifié; puis un processus d'identification d'un modèle plus simple est entrepris. La logique formelle de vérification n'est pas entièrement respectée puisque le modèle théorique de départ est déjà supposé modifié avant le test crucial. Cependant, comme les informations disponibles sont prises en considération

à chacun des stages du processus d'analyse, la nécessité de vérifier le modèle théorique comme élément fondamental d'un processus de recherche est maintenue.

Dans l'état actuel du développement de l'analyse de données par les modèles log-linéaires, le programme que nous présentons permet d'aborder de façon compréhensive l'étude des tableaux de contingence avec ses raffinements les plus importants. En une seule séance de travail et à très peu de frais, en conversant avec l'ordinateur, mais sous l'entier contrôle de l'utilisateur, le chercheur peut soumettre ses hypothèses à l'analyse statistique ou procéder à l'exploration systématique de ses données. En effet, le programme ne contient aucune procédure automatique de détection d'interaction. L'utilisateur est entièrement maître du déroulement du programme et de l'application des sous-routines. Il n'est limité que par ce qu'il ne contient pas.

## I. Exemples d'utilisation

Les données (tableau 1) utilisées dans cet exemple, proviennent de la cohorte anglophone, secondaire V, temps I, de l'enquête A.S.O.P.E. (groupe d'étude sur les aspirations scolaires et les orientations professionnelles des étudiants du Québec). Les variables dénommées sous les caractères C, S, P, A sont respectivement les suivantes: le statut socio-économique mesuré par l'éducation du père, le sexe de l'étudiant, le programme d'études dans lequel il est inscrit et ses aspirations professionnelles. Toutes

les variables ont été dichotomisées pour augmenter la simplicité de l'exposé; la signification des catégories de chaque variable est la suivante: $C_1$: SSE élevé, $C_2$: SEE bas; $S_1$: masculin, $S_2$: féminin; $P_1$: programme général, $P_2$: programme professionnel; $A_1$: aspirations professionnelles élevées, $A_2$: aspirations professionnelles basses.

Nous chercherons à valider l'hypothèse suivante: les aspirations professionnelles et la fréquentation d'un programme d'études sont le reflet des éléments de définition de la division du travail dans une société. Etant donné les éléments social (variable C) et sexuel (variable S) de la division du travail, nous nous attendrons à ce que les paramètres CSA (la relation statut socio-économique - sexe - aspirations professionnelles) et CSP (la relation statut socio-économique - sexe - programmes professionnels) soient significatifs et inclus dans le modèle théorique. D'autre part, la relation PA (programme professionnel - aspirations professionnelles) existe certainement quoiqu'elle ne soit pas de prime intérêt ici. En somme, nous chercherons à valider les variantes d'un modèle qui peut s'écrire PA, CSA, CSP.

A) *Les coefficients d'association*

Dans un premier temps des coefficients d'association seront estimés. Ils permettent une vérification préliminaire des hypothèses de départ.

Ces coefficients dénommés Tau, Lambda et Lambda standardisé estiment les chances qu'une des catégories d'une variable soit reliée

à l'une des catégories d'une autre ou de plusieurs variables[1]

Par exemple étant donné les variables dichotomiques α et β nous obtenons le tableau de contingence suivant:

| | | β = 1 | β = 2 | |
|---|---|---|---|---|
| α | 1 | $f_{11}$ | $f_{12}$ | $f_{1.}$ |
| | 2 | $f_{21}$ | $f_{22}$ | $f_{2.}$ |
| | | $f_{.1}$ | $f_{.2}$ | $f_{..}$ |

Figure 1: Tableau de contingence des fréquences associées à α et β.

où les f sont les fréquences accompagnées des indices appropriés aux cellules ou marginales du tableau. Les sommes suivantes peuvent être définies:

$$(1) \qquad f_{1.} = f_{11} + f_{12}$$

$$f_{2.} = f_{21} + f_{22}$$

$$f_{.1} = f_{11} + f_{21}$$

$$f_{.2} = f_{12} + f_{22}$$

$$f_{..} = f_{11} + f_{12} + f_{21} + f_{22} = f_{1.} + f_{2.} = f_{.1} + f_{.2}$$

---

1. Les modèles log-linéaires permettent de traiter des variables polytomiques. Cependant le nombre de coefficients à évaluer devient rapidement aberrant de telle sorte qu'ils ne sont utilisables dans ce cas qu'avec un très petit nombre de variables. Par exemple, un système de trois variables quadritomiques I, J, K produira (I-1) + (J-1) + (K-1) + (I-1)(J-1) + (I-1)(J-1)(K-1) termes, soit 63 coefficients. Ce qui est inévitable dans le cas de données nominales.

La méthode utilise deux coefficients d'association de base: Tau et Lambda (L.A. Goodman, 1972b). Tous les autres coefficients en sont des modifications. Les rapports entre Tau et Lambda sont simples comme nous le verrons.

Le coefficient Tau, dans le cas où deux variables dichotomiques sont utilisées, n'est rien d'autre qu'une fonction du produit croisé de la table de contingence de la figure 1:

$$(2) \qquad \tau_{11}^{\alpha\beta} = \left(\frac{f_{11} \times f_{22}}{f_{21} \times f_{12}}\right)^{\frac{1}{4}}$$

Si $(f_{11} \times f_{22}) = (f_{21} \times f_{12})$ alors, $\tau_{11}^{\alpha\beta} = 1$, d'où il n'y a pas de relation entre $\alpha$ et $\beta$. De même il est possible de calculer les chances qu'un individu soit dans la catégorie 1 de $\alpha$ plutôt que dans la catégorie 2 de $\alpha$ quand $\beta$ est maintenu constant:

$$(3) \qquad \tau_{1}^{\alpha} = \left(\frac{f_{11} \times f_{12}}{f_{21} \times f_{22}}\right)^{\frac{1}{4}}$$

Notez que les premiers indices des f du numérateur (catégorie 1 de $\alpha$) et du dénominateur (catégorie 2 de $\alpha$) sont constants. S'il n'y avait qu'une seule variable dichotomique $\alpha$, Tau de $\alpha$ serait égal à:

$$(4) \qquad \tau_{1}^{\alpha} = \left(\frac{f_1}{f_2}\right)^{\frac{1}{2}}$$

L'exposant de la formule de Tau est un facteur de pondération qu'il sera aisé de comprendre lorsque nous aborderons le calcul des coefficients Lamdba.

Par symétrie avec 3:

$$(5) \qquad \tau_1^{\beta} = \frac{f_{11} \times f_{21}}{f_{12} \times f_{22}}^{\frac{1}{4}}$$

$\tau_1^{\alpha}$ est donc une mesure des chances d'un individu d'appartenir à $\alpha_1$ plutôt qu'à $\alpha_2$ et $\tau_1^{\beta}$, une mesure des chances d'un individu d'appartenir à $\beta_1$ plutôt qu'à $\beta_2$ dans un tableau bivarié. $\tau_{11}^{\alpha\beta}$ peut-être considéré comme une mesure des chances d'un individu d'appartenir à $\alpha_1\beta_1$ plutôt qu'à $\alpha_1\beta_1$.

Notons que:

$$(6) \quad \tau_2^{\alpha} = \frac{f_{21} \times f_{22}}{f_{11} \times f_{12}} = 1/\tau_1^{\alpha}$$

$$\tau_2^{\beta} = \frac{f_{12} \times f_{22}}{f_{11} \times f_{21}} = 1/\tau_2^{\beta}$$

$$\tau_{12}^{\alpha\beta} = \frac{f_{12} \times f_{21}}{f_{22} \times f_{11}} = \tau_{21}^{\alpha\beta} = 1/\tau_{11}^{\alpha\beta} = 1/\tau_{22}^{\alpha\beta}$$

d'où, dans le cas dichotomique, seuls les coefficients $\tau_1^{\alpha}$, $\tau_1^{\beta}$ et $\tau_{11}^{\alpha\beta}$ sont nécessaires pour évaluer l'ensemble des coefficients générés par la table de contingence. Dans le cas polytomique le nombre de coefficients nécessaires dépend du nombre de degrés de liberté du tableau de contingence. L'extension de cette interprétation du coefficient $\tau$ au cas général à N variables polytomiques est directe.

Les Lambda sont comme les Tau une fonction de probabilité. A partir du tableau de contingence de la figure 1, nous obtenons le tableau des probabilités suivant:

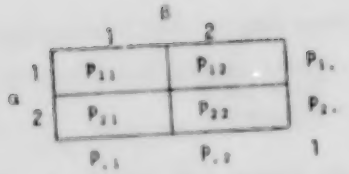


Figure 2: Tableau de contingence des probabilités associées à α et β lorsque:

$$p_{11} = \frac{f_{11}}{f_{..}}, \text{ etc.}$$

Etant donné:

$$(7) \qquad v_{ij} = \log p_{ij}$$

$$v_{i.} = \Sigma_j v_{ij}/J$$

Alors,

$$(8) \qquad v_{.j} = \Sigma_i v_{ij}/I$$

$$v_{..} = \Sigma_{i,j} v_{ij}/IJ$$

où I et J tiennent lieu du nombre de catégories de chacune des variables et i = 1...I et j = 1...J. Pour le tableau de la figure 2:

$$(9) \qquad v_{1.} = v_{11} + v_{12}/2$$

$$v_{.1} = v_{21} + v_{11}/2$$

$$v_{..} = v_{11} + v_{12} + v_{21} + v_{22}/2x2$$

Les lambda sont des fonctions de probabilités, des rapports de proportions. Ainsi,

$$(10) \qquad \lambda_1^{\alpha} = v_{1.} - v_{..}$$

$$\lambda_1^{\beta} = v_{.1} - v_{..}$$

$$\lambda_{11}^{\alpha\beta} = v_{11} - v_{.1} - v_{1.} + v_{..}$$

Lambda de $\alpha_i$ suppose connues les marginales correspondant à la fonction de probabilité $v_{ij}$, c'est-à-dire en vertu de (7), (8) et (9):

$$(11) \qquad v_{i\cdot} = \frac{v_{i1} + v_{i2}}{2}$$

$$v_{\cdot\cdot} = \frac{v_{11} + v_{12} + v_{12} + v_{22}}{4}$$

d'où:

$$(12) \qquad \lambda_i^{\alpha} = \frac{v_{i1} + v_{i2}}{2} - \frac{v_{11} + v_{21} + v_{12} + v_{22}}{4}$$

$$\lambda_i^{\alpha} = \frac{v_{i1} + v_{i2} + v_{i1} + v_{i2}}{4}$$

$$\lambda_i^{\alpha} = \frac{(\log p_{i1} + \log p_{i2} + \log p_{i1} + \log p_{i2})}{4}$$

$$\lambda_i^{\alpha} = \log \left[ \frac{p_{i1} \times p_{i2}}{p_{i1} \times p_{i2}} \right]^{\frac{1}{4}}$$

$$\lambda_i^{\alpha} = \log \left[ \frac{f_{i1}/f_{\cdot\cdot} \times f_{i2}/f_{\cdot\cdot}}{f_{i1}/f_{\cdot\cdot} \times f_{i1}/f_{\cdot\cdot}} \right]^{\frac{1}{4}}$$

$$\lambda_i^{\alpha} = \log \left[ \frac{f_{i1} \times f_{i2}}{f_{i1} \times f_{i2}} \right]^{\frac{1}{4}}$$

D'où en vertu de (3):

$$(13) \qquad \lambda_i^{\alpha} = \log \tau_i^{\alpha}$$

L'argument est symétrique pour $\lambda_{11}^{\beta}$, et $\lambda_{11}^{\alpha\beta}$, et directement applicable au cas de N variables polytomiques. Puisque Lambda est le logarithme de Tau, Lambda sera égal à 0 lorsque $\tau = 1$, négatif lorsque $\tau < 1$, et positif quand $\tau > 1$. En somme, il est d'interprétation plus aisée et se comporte à cet égard comme le $\phi$ ou le r de Pearsons.

Bien entendu tout comme (6):

$$(14) \qquad \lambda_1^{\alpha} = -\lambda_2^{\alpha} \; ; \qquad \lambda_1^{\beta} = -\lambda_2^{\beta}$$

$$\lambda_{11}^{\alpha\beta} = \lambda_{22}^{\alpha\beta} = -\lambda_{21}^{\alpha\beta} = -\lambda_{12}^{\alpha\beta}$$

Quant aux Lambda standardisés, ils sont facilement interprétables et se prêtent à un premier test de significativité des relations entre variables prises une à une[1]. Soit un $\lambda$ égal à zéro (0), sa moyenne échantillonnale est égale à zéro et sa variance à l'unité ($\sigma^2 = 1$). En conséquence, pour qu'un $\lambda$ soit égal à zéro dans un intervalle de confiance de $\alpha = .05$, il suffit que le $\lambda$ standardisé, dont la variance est l'unité, soit compris dans l'intervalle de confiance correspondant à $\alpha$ dans une table de test $\tau$ (L.A. Goodman, 1971). Etant donné un échantillon égal ou supérieur à 120, l'intervalle est de $\pm 1.96$. Puisque le chercheur est habituellement intéressé à un modèle qui contient plusieurs paramètres relationnels, une correction s'impose. Nous négligerons ici le détail de cette procédure (L.A. Goodman, 1970).

---

1. La table $\tau$ est utilisée puisque la moyenne de la population est inconnue.

Dans ce qui va suivre nous nous appliquerons à l'étude de l'hypothèse de départ à l'aide de l'instrument d'analyse préliminaire que sont les coefficients Lambda et Lambda standardisé.

La lecture du tableau 2 montre, étant donné le seul de 1.96, une distribution inégalitaire de nos individus sur deux des quatre variables retenues: a) Lambda de $C_1(\lambda_1^C)$: il y a beaucoup plus d'enfants de parents de SSE bas que d'enfants de parents de SSE élevé au secondaire V, anglophone; b) $S_1(\lambda_1^S)$: il y a autant de filles que de garçons; c) Lambda de $P_1(\lambda_1^P)$: plus d'étudiants au programme général qu'au programme professionnel et d) Lambda $A_1(\lambda_1^A)$: autant d'étudiants aux aspirations professionnelles basses qu'élevées. Les relations d'ordre 1 qualifient ces conclusions: a) Lambda de $C_1S_1(\lambda_{11}^{CS})$ les étudiants proviennent du même milieu socio-économique que les étudiantes, b) Lambda de $C_1P_1(\lambda_{11}^{CP})$: plus le SSE est élevé plus il y a de chances qu'un individu soit inscrit au programme général et c) Lamdba de $C_1A_1(\lambda_{11}^{CA})$: plus ses aspirations professionnelles sont élevées;[1] d) Lambda de $S_1P_1(\lambda_{11}^{SP})$: les étudiants et les étudiantes fréquentent un même niveau de programme, d'autre part e) Lambda $S_1A_1(\lambda_{11}^{SA})$: les étudiantes expriment des aspirations professionnelles de même niveau que les étudiants, enfin f) Lambda de $P_1A_1(\lambda_{11}^{PA})$: les aspirations professionnelles et le programme d'études sont étroitement liés.

---

1. En conséquence, si le secondaire V apparaissait à la lumière de l'analyse des paramètres des moyennes $\lambda^S$, $\lambda^P$, $\lambda^C$ et $\lambda^A$ comme un instrument démocratique d'accès à l'éducation, l'étude des paramètres d'interaction d'ordre un limite singulièrement cette proposition.

Tableau 1

Distribution de fréquences d'un échantillon d'étudiants anglophones de secondaire V du Québec

| Statut socio-économique (C)[1] | Sexe de l'étudiant (S) | Programme fréquenté (P) | Aspirations professionnelles A 1 | A 2 |
|---|---|---|---|---|
| 1 | 1 | 1 | 318 | 91 |
| 1 | 1 | 2 | 10 | 14 |
| 1 | 2 | 1 | 335 | 97 |
| 1 | 2 | 2 | 13 | 12 |
| 2 | 1 | 1 | 359 | 229 |
| 2 | 1 | 2 | 14 | 59 |
| 2 | 2 | 1 | 323 | 229 |
| 2 | 2 | 2 | 33 | 82 |

1. Voir dans le texte la signification des lettres et indices.

Pour avoir une image plus complète il semble que les paramètres relationnels $C_1S_1P_1$ et $S_1P_1A_1$ ($\lambda^{CSP}_{1\,1\,1}$ et $\lambda^{SPA}_{1\,1\,1}$) puissent être ajoutés à cette discussion, étant donné que la valeur de leurs Lambda standardisés soit près de ±1.96. Leur contribution pourrait être interprétée de la façon suivante: les relations $C_1S_1$ et $S_1A_1$ dépendent des catégories de P. Le modèle qui résulte de cette discussion est le suivant: CA, CSP, SPA. (Nous n'indiquerons plus les indices des variables puisque, étant dichotomiques, il ne peut y avoir d'ambiguïté).

Tableau 2

Tau, Lambda et Lambda standardisé du modèle non-saturé

| Paramètres relationnels | Tau | Lambda | Lambda standardisé |
|---|---|---|---|
| $C_1$ | 0.6664 | -0.4059 | - 9.2984 |
| $S_1$ | 0.9217 | -0.0816 | - 1.8684 |
| $P_1$ | 3.1937 | 1.1612 | 26.5979 |
| $A_1$ | 1.0434 | 0.0424 | 0.9721 |
| $C_1S_1$ | 1.0550 | 0.0536 | 1.2266 |
| $C_1P_1$ | 1.1884 | 0.1726 | 3.9536 |
| $C_1A_1$ | 1.2672 | 0.2368 | 5.4249 |
| $S_1P_1$ | 1.0836 | 0.0803 | 1.8390 |
| $S_1A_1$ | 0.9496 | -0.0517 | - 1.1844 |
| $P_1A_1$ | 1.4450 | 0.3681 | 8.4312 |
| $C_1S_1P_1$ | 0.9220 | -0.0813 | - 1.8611 |
| $C_1S_1A_1$ | 1.0011 | 0.0011 | 0.0256 |
| $C_1P_1A_1$ | 0.9756 | -0.0247 | - 0.5657 |
| $S_1P_1A_1$ | 1.0686 | 0.0664 | 1.5207 |
| $C_1S_1P_1A_1$ | 0.9872 | -0.0129 | - 0.2944 |

Ce modèle est hiérarchique, concrètement: les informations données par les paramètres relationnels d'ordre N augmentent d'un degré celles fournies par les paramètres relationnels d'ordre N-1. Le sens de l'expression "analyse hiérarchique" est contenu dans cette explication. Ainsi, les paramètres CS, CP, SP sont inclus dans le paramètre CSP.

Théoriquement le modèle CA, CSP, SPA peut s'exprimer ainsi: la distribution de fréquences dans le tableau dichotomique à quatre entrées n'est pas uniforme selon que les individus sont inscrits au programme général plutôt qu'au programme professionnel, mais il existe aussi des différences entre les aspirations professionnelles des étudiants établies en fonction de leur statut socio-économique d'origine et de leur sexe. Le modèle vérifie la validité d'une proposition théorique qui aurait posée une structuration différentielle de l'univers scolaire en fonction du programme d'étude et des éléments de division du travail que sont le sexe et le statut social d'origine. Mais contrairement à notre hypothèse de départ, qui privilégiait ces deux dernières variables, la variable primordiale apparaît être le programme d'étude qui semble lié à la fois à la classe sociale d'appartenance, à la distribution des étudiants selon leur sexe et aux aspirations professionnelles. La différence majeure entre le modèle théorique de départ et le modèle corrigé est l'absence du terme CSA, qui indique une interaction entre classe sociale, sexe et aspiration. Le terme SPA peut être considéré comme une spécification du terme PA. De même la classe sociale joue-t-elle directement sur les aspirations professionnelles, sans aucun intermédiaire. Un nouveau terme nous est cependant proposé, SPA, qui replace en perspective l'effet du sexe: soit un effet indépendant, autonome de la classe sociale lorsque sa relation avec les aspirations (A) est étudiée. Reste à vérifier la validité de ce modèle, puisqu'il ne tient qu'à une analyse approximative des lambda standardisés, pour laquelle nous avons ac-

cepté un degré de significativité individuelle plus bas que 5% comptant sur le fait que leur introduction dans un modèle requiert habituellement un ajustement de l'intervalle de confiance des lambda standardisés vers le bas.

### iii Les modèles

Les lambda obtenus à partir des fréquences observées décrivent un modèle de relation qui évalue exactement ces fréquences. Les lambda sont donc des paramètres qui peuvent être conçus comme des effets spécifiques dont les fréquences observées sont fonction. Ainsi, l'équation 15 prédit les fréquences observées d'une table de contingence à quatre entrées. Elle est l'équation du modèle saturé (i.e. toutes les relations sont incluses dans le modèle):

$$(15)\quad \ln F^{CSPA}_{ijkl} = \theta + \lambda^{C}_{i} + \lambda^{S}_{j} + \lambda^{P}_{k} + \lambda^{A}_{l} + \lambda^{CS}_{ij} + \lambda^{CP}_{ik} + \lambda^{CA}_{il} + \lambda^{SP}_{jk} + \lambda^{SA}_{jl} + \lambda^{PA}_{kl} + \lambda^{CSP}_{ijk} + \lambda^{CSA}_{ijl} + \lambda^{CPA}_{ikl} + \lambda^{SPA}_{jkl} + \lambda^{CSPA}_{ijkl}$$

où $F^{CSPA}_{ijkl}$ est la fréquence du tableau de contingence des variables C, S, P, A de catégories i, j, k et l. $\theta$, est une constante qui adapte la somme des coefficients $\lambda$ au nombre N d'individus dans l'échantillon, tandis que les indices i, j, k et l des variables C, S, P et A respectivement indiquent les rapports mesurés quant $i = 1, \ldots, I-1$, $j = 1, \ldots, J-1$, $k = 1, \ldots, K-1$, $l = 1, \ldots, L-1$ (I, J, K et L sont les nombres de catégories des variables C, S, P et A).

Mais le chercheur n'est pas intéressé par le modèle saturé. Impliquant l'ensemble des relations, il suppose qu'un ensemble lourd d'informations est nécessaire pour reproduire un phénomène décrit par un tableau de fréquences. En fait, il doit être considéré comme impraticable. Ce qu'il est en réalité quand 4 variables ou plus composent l'univers expérimental. L'équation (non-saturée) 16 décrit l'hypothèse que nous avions retenue au départ:

$$(16)\quad \ln F^{CSPA}_{ijkl} = \theta + \lambda^{C}_{i} + \lambda^{S}_{j} + \lambda^{P}_{k} + \lambda^{A}_{l} + \lambda^{CS}_{ij} + \lambda^{CP}_{ik} + \lambda^{CA}_{il} + \lambda^{SP}_{jk} + \lambda^{SA}_{jl} + \lambda^{PA}_{kl} + \lambda^{CSP}_{ijk} + \lambda^{CSA}_{ijl}$$

où $F^{CSPA}_{ijkl}$ tient lieu de fréquences estimées par le modèle. Ce modèle s'écrit (PA, CSP, CSA). Les autres termes de l'équation sont déjà familiers. L'équation est donc une traduction formelle d'un modèle théorique. Nous procéderons à l'analyse des modèles théoriques précédemment dégagés de l'étude des $\lambda$. Les fréquences théoriques obtenues à l'aide de la traduction formelle des modèles nous servirons à valider les modèles théoriques en utilisant le test de Khi deux bien connu où le test du Khi deux du maximum de vraisemblance dit $G^2$ (Y.M.M. Bishop et al, 1975).

Les tableaux 3 et 4 illustrent les résultats obtenus par le programme d'évaluation des fréquences estimées, soit les fréquences de divers modèles évalués (tableau 4) et leur $G^2$ correspondants. L'utilisateur est intéressé à obtenir la meilleure approximation de ces données d'origine par son modèle, tout en respectant les règles d'une

grande économie. En fixant $\alpha = .05$, le modèle constitue l'hypothèse nulle dans le cadre du test $X^2$ bien connu. Ici, avec trois degrés de liberté (les degrés deliberté sont égaux à la somme des relations d'un modèle saturé moins la somme des relations égales à zéro) et un $G^2$ de 3.5, le modèle théorique (modèle 3) ne peut être rejeté, ($.50<P<.75$ selon la table de $X^2$). Cependant, le modèle modifié, le modèle PA, CSA, CSP (modèle 5), reproduit mieux les fréquences observées. En conséquence, la théorie qui a servi à créer le modèle ne peut être rejetée, mais des modèles alternatifs peuvent être proposés.

Si plusieurs modèles peuvent reproduire les fréquences observées, peut-être existe-il un modèle minimum statistiquement satisfaisant. Nous sommes donc à la recherche d'un modèle parcimonieux.

Le modèle CP, CA, PA qui exclut la variable sexe du modèle et qui ne concerne que les interactions d'ordre 1 entre C, P et A reproduit les fréquences observées avec un $G^2$ de 14.89 et 9 degrés de liberté. Autrement dit l'hypothèse d'hétérogénéité entre les fréquences théoriques et les fréquences observées doit être rejetée. Cette hypothèse rejette fortement le sexe comme un critère de division sociale de l'orientation professionnelle, pour ne conserver que la classe sociale, le programme d'études fréquenté et les aspirations scolaires.

Pour des raisons théoriques, le chercheur peut accepter le modèle CA, CSP, SPA (soit le modèle $M_5$ du tableau 4). Mais avant

de se faire, il vaut mieux s'assurer de la validité de l'apport de chacun des paramètres critiques de ce modèle soit SPA et CSP. L'analyse de la baisse de la variation non-expliquée par $M_5$ permet de nous livrer à ce travail. Le tableau 5 montre que ni $\lambda_{CSP}$, ni $\lambda_{SPA}$ ne sont significatifs à .05 selon la table de Khi deux, quoique l'un et l'autre soit bien près de la valeur critique de 3.84. De telle sorte que, pour des raisons théoriques extérieures aux raisons statistiques, il sera justifiable de relâcher le seuil de significativité de .05.

Tableau 3

Fréquences estimées par les modèles C, S, P, A; CA, CSP, SPA, et CP, CA, PA

| Modèle vérifié: C, S, P, A | | | | | Modèle vérifié: CA, CSP, SPA | | | | | Modèle vérifié: CP, CA, PA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | S | P | A 1 | A 2 | C | S | P | A 1 | A 2 | C | S | P | A 1 | A 2 |
| 1 | 1 | 1 | 248.36 | 143.71 | 1 | 1 | 1 | 320.98 | 88.02 | 1 | 1 | 1 | 327.03 | 93.47 |
| 1 | 1 | 2 | 29.71 | 17.91 | 1 | 1 | 2 | 9.09 | 14.91 | 1 | 1 | 2 | 10.97 | 13.53 |
| 1 | 2 | 1 | 255.17 | 147.65 | 1 | 2 | 1 | 333.55 | 98.45 | 1 | 2 | 1 | 327.03 | 93.47 |
| 1 | 2 | 2 | 30.54 | 17.66 | 1 | 2 | 2 | 12.38 | 12.62 | 1 | 2 | 2 | 10.97 | 13.53 |
| 2 | 1 | 1 | 370.59 | 214.44 | 2 | 1 | 1 | 356.02 | 231.98 | 2 | 1 | 1 | 340.47 | 229.53 |
| 2 | 1 | 2 | 44.34 | 25.65 | 2 | 1 | 2 | 14.91 | 58.09 | 2 | 1 | 2 | 24.03 | 69.97 |
| 2 | 2 | 1 | 380.75 | 220.32 | 2 | 2 | 1 | 324.45 | 227.55 | 2 | 2 | 2 | 340.47 | 229.53 |
| 2 | 2 | 2 | 45.55 | 26.36 | 2 | 2 | 2 | 33.62 | 81.38 | 2 | 2 | 2 | 24.03 | 69.97 |

Un analyste strict ne serait cependant pas satisfait de ce résultat et pousserait l'étude plus loin. Le plus petit modèle significatif obtenu est le modèle $M_9$. Pour des raisons théoriques le chercheur peut être amené à le préférer. D'autant plus qu'il résiste beaucoup mieux au critère de .05 préalablement fixé. Le tableau 6 l'illustre parfaitement, où les $\lambda_{CP}$, $\lambda_{CA}$ et $\lambda_{PA}$ sont tous significatifs à .05.

Tableau 4

Modèles vérifiés, leurs degrés de liberté et leurs Khi deux du rapport de vraisemblance

| Modèles | Degrés de liberté | Khi deux du rapport de vraisemblance |
|---|---|---|
| (1) PA, CSA, CSP | 3 | 3.50 |
| (2) CSA, CSP | 4 | 103.33 |
| (3) PA, CSA, CP, SP | 4 | 5.73 |
| (4) PA, CSP, CA, SA | 4 | 3.60 |
| (5) CA, CSP, SPA | 3 | .53 |
| (6) CA, CS, CP, SPA | 4 | 3.99 |
| (7) CS, CP, CA, SP, SA, PA | 6 | 6.10 |
| (8) CA, CP, PA, S | 10 | 14.49 |
| (9) CA, CP, PA | 9 | 14.84 |
| (10) CA, CP | 10 | 115.85 |
| (11) CP, PA | 10 | 80.86 |
| (12) CA, PA | 10 | 35.41 |
| (13) C, S, P, A | 9 | 265.44 |
| (14) Hypothèse d'équiprobabilité | 14 | 2079.92 |

Confronté à un tel résultat, le chercheur préférera le modèle le plus simple. On peut en effet considérer que le modèle saturé CSPA qui contient l'ensemble des $\lambda$ est entièrement au hasard puisqu'il suppose la connaissance de tous les éléments d'informations disponibles pour se reproduire lui-même. Ainsi le vecteur 1001101110101101 reproduit les résultats d'une séquence de tirs d'une pièce de monnaie (pile: 1; face: 0) et seule la connaissance de toute la séquence permet de la connaître (G.J. Chatin, 1975). Le vecteur 1001011010010110 n'est pas au hasard. La connaissance de 10010110, soit la moitié des éléments d'informations est nécessaire pour le reproduire; de plus il serait facile d'édicter des règles pour se faire. En conséquence, il est permis d'écrire que plus d'éléments sont nécessaires pour reproduire un vecteur quelconque, plus le vecteur est construit au hasard. Il devient clair que le chercheur préférera le modèle le plus petit étant donné certaines règles (c'est-à-dire un seuil de significativité par exemple). En ce cas, le modèle 9 du tableau 4 doit être préféré puisqu'il s'écrit:

$$(17) \quad \ln F^{CSPA}_{ijkl} = \theta + \lambda^{C}_{i} + \lambda^{P}_{k} + \lambda^{A}_{l} + \lambda^{CP}_{ik} + \lambda^{CA}_{il} + \lambda^{PA}_{kl}$$

et exige sept éléments d'information, tandis que $M_5$ en exige 13 (voir équation 16). Le choix du chercheur se portera donc sur l'équation 17.

Tableau 5

Analyse de la variation attribuée à $\lambda_{CPA}$ et $\lambda_{SPA}$ dans $M_5$

| | Modèles | DL | Khi deux du rapport de vraisemblance |
|---|---|---|---|
| (1) | Variation due à la dépendance entre C, P, S, A non expliquée par $M_6$ | 4 | 3.99 |
| (1a) | Variation due à la dépendance entre C, P, S, A non-expliquée par $M_5$ | 3 | .53 |
| (1b)* | Variation expliquée par $\lambda_{CSP}$ dans $M_5$ | 1 | 3.46 |
| (2) | Variation due à la dépendance entre C, P, S, A non expliquée par $M_4$ | 4 | 3.60 |
| (2b) | Variation expliquée par $\lambda_{SPA}$ dans $M_5$ | 1 | 3.07 |

* La ligne 1b donne la variation expliquée par le paramètre CSP dans $M_5$ puisque le modèle $M_6$ contient tous les paramètres de $M_5$ sauf CSP.

Le tableau 6 peut se traduire en coefficients standards, comparables entre eux, qui permettent une lecture plus rapide des résultats. Les coefficients obtenus s'interprètent 1o) comme le coefficient de détermination multiple, 2o) le coefficient de détermination multi-partielle (partition de 1 dans le tableau 6 en 1a et 1b) et 3o) le coefficient de détermination partielle (partition de (1a) dans le tableau 6) de l'analyse de régression multivariée ordinaire.

Tableau 6

Analyse de la variation attribuée à $\lambda_{SP}$, $\lambda_{CA}$ et $\lambda_{PA}$ dans $M_9$

| | Modèle | Degrés de liberté | Khi deux du rapport de vraisemblance |
|---|---|---|---|
| (1) | Variation due à C, S, P, A | 11 | 265.44* |
| (1a) | Variation due à C, S, P, A non expliquée par $M_9$ | 9 | 14.39* |
| (1b) | Variation expliquée par $M_9$ | 2 | 240.55* |
| (1a.1a) | Variation due à C, S, P, A non expliquée par $M_{10}$ | 10 | 115.85* |
| (1a.2a) | Variation expliquée par $\lambda_{PA}$ dans $M_9$ | 1 | 100.94* |
| (1a.2b) | Variation due à C, S, P, A non expliquée par $M_{11}$ | 10 | 94.95* |
| (1a.2b) | Variation expliquée par $\lambda_{CA}$ dans $M_9$ | 1 | 80.86* |
| (1a.1c) | Variation due à C, S, P, A non expliquée par $M_{12}$ | 10 | 35.41* |
| (1a.2c) | Variation expliquée par $\lambda_{CP}$ | | |

* Significatif à .05.

Il apparaît (tableau 7) que le modèle $M_9$ produit une baisse de 98% de la variation totale non-expliquée (coefficient de détermination multiple calculé en tenant compte de l'hypothèse d'équiprobabilité considérée comme celle portant le maximum de variation dans la table de fréquences). D'autre part, le coefficient de détermination multi-partielle révèle un gain de 93% quand on retient $M_9$ plutôt que l'hypothèse d'indépendance.

Les poids répartis de chacun des paramètres sont donnés par les coefficients de détermination partielle. Il apparaît que $\lambda_{CP}$ soit le moins important dans le modèle $M_9$ tandis que $\lambda_{CA}$ et $\lambda_{PA}$ sont approximativement d'importance égale. Autrement dit, dans un système de quatre variables C, S, P, A, la structure de la division sociale de l'orientation professionnelle se décrit à l'aide du rapport privilégié entre l'appartenance de classe et les aspirations professionnelles d'une part, le programme scolaire et les aspirations professionnelles d'autre part. Le rapport entre la classe sociale d'origine et le programme d'études ne vient qu'après, tandis que le sexe de l'étudiant est complètement exclus du modèle (le signe négatif du coefficient de détermination partielle de $\lambda_S$ s'explique par la pondération des $G^2$ de $M_8$ et $M_9$ par leurs degrés de liberté respectifs).

Puisque $M_9$ est satisfaisant, économique et d'interprétation aisée, nous considérerons sa distribution de fréquences pour fin d'analyse. Le tableau 8 évalue l'ensemble des $\lambda$. Ceux qui ont une valeur nulle sont exclus du modèle, (une erreur d'évaluation est illustrée par la valeur de $\lambda^{CPA}_{111} = -.0001$, valeur assez petite pour ne pas

en tenir compte). Les $\lambda$ non nuls sont pratiquement égaux à leurs correspondants du tableau 3, ce qui indique une bonne fiabilité des fréquences estimées. Remarquons que $\lambda_1^A$ est évalué par le modèle malgré le fait qu'il ne soit pas significatif puisque $\lambda_1^A$ est contenu dans CA et PA. En résumé $M_9$ permet d'écrire: 1o que le niveau secondaire V anglophone est plus fréquenté par des étudiants de classe sociale d'origine basse, 2o plus d'étudiants sont inscrits au programme général qu'au programme professionnel, 3o les étudiants d'origine sociale élevée ($C_1$) ont plus de chances de s'inscrire au programme général qui s'ouvre sur le cégep que les étudiants d'origine sociale basse ($C_2$), 4o les aspirations professionnelles des $C_1$ sont plus élevées que celles des $C_2$ et 5o la fréquentation du programme général est en rapport avec l'expression d'aspirations professionnelles élevées. La lecture du tableau 9 montre clairement que le sexe n'est en rapport avec aucune variable et n'a pas d'effet de par la simple distribution des étudiants selon leur sexe (voir $\lambda_1^S$).

Tableau 7

Les coefficients de détermination multiple, multipartielle et partielle

| | Valeurs de $R^2$ |
|---|---|
| $R^2_{f \cdot M_9}$ | .988 |
| $R^2_{f(M_9) \cdot \lambda^C \lambda^S \lambda^P \lambda^A}$ | .931 |
| $R^2_{f\lambda PA \cdot M_{10}}$ | .857 |
| $R^2_{f\lambda CA \cdot M_{11}}$ | .825 |
| $R^2_{f\lambda CP \cdot M_{12}}$ | .532 |
| $R^2_{f\lambda S \cdot M_9}$ | -.097 |

f désigne les fréquences observées

Tableau 8

Evaluation des λ selon les fréquences estimées au modèle 9

| Paramètres relationnels | Tau | Lambda | Lambda standardisés |
|---|---|---|---|
| $C_1$ | 0.6566 | -0.4207 | -9.8323 |
| $S_1$ | 1.0000 | 0.0000 | 0.0000 |
| $P_1$ | 3.1450 | 1.1458 | 26.2264 |
| $A_1$ | 1.0471 | 0.0460 | 1.0755 |
| $C_1S_1$ | 1.0000 | 0.0000 | 0.0000 |
| $C_1P_1$ | 1.2045 | 0.1861 | 4.3484 |
| $C_1A_1$ | 1.2394 | 0.2146 | 5.0159 |
| $S_1P_1$ | 1.0000 | 0.0000 | 0.0000 |
| $S_1A_1$ | 1.0000 | 0.0000 | 0.0000 |
| $P_1A_1$ | 1.4415 | 0.3657 | 8.5448 |
| $C_1S_1P_1$ | 1.0000 | 0.0000 | 0.0000 |
| $C_1S_1A_1$ | 1.0000 | 0.0000 | 0.0000 |
| $C_1P_1A_1$ | 0.9999 | -0.0001 | -0.0026 |
| $S_1P_1A_1$ | 1.0000 | 0.0000 | 0.0000 |
| $C_1S_1P_1A_1$ | 1.0000 | 0.0000 | 0.0000 |

C) *Une lecture plus concrète du modèle: la normalisation des fréquences estimées*

La signification des valeurs des $\lambda$ demeure cependant obscure. Etant une moyenne géométrique de proportion, le sens nous en échappe. Il manque une interprétation "physique" immédiatement préhensible à l'entendement.

La procédure itérative de F. Mosteller (1968) nous donnera l'occasion d'interpréter facilement les $\lambda$ et de les convertir en chances individuelles.

$M_9$ peut se traduire en trois tables de fréquences suffisantes pour décrire l'ensemble du phénomène sous observation. Les trois tables obtenues se prêtent bien à l'exécution de la procédure de normalisation de Mosteller. Ces trois tables sont exactement évaluées par les paramètres du modèle CP, CA, PA du tableau B. L'utilisation des fréquences observées ou des fréquences théoriques produirait un résultat identique. Le modèle théorique reçoit donc une interprétation physique simple: ses paramètres établissent la liste des tableaux de contingence nécessaires à l'étude de l'univers expérimental clos constitué des fréquences observées.

La procédure de normalisation des fréquences annule l'effet de l'inégalité de distribution des événements dans chacune des variables et isole l'effet d'interaction. Dans le cas qui nous préoccupe, 69.0% des étudiants ont des aspirations professionnelles comformes au programme dans lequel ils sont inscrits ($P_{11}^{PA} + P_{22}^{PA}$ où

P indique la proportion d'étudiants dans chacune des cases de la table de proportionalité PA). De même 62.7% des étudiants se dirigent vers le même type de niveau de cours que leur classe d'origine (i.e. programme professionnel - classe d'origine à statut peu élevé, programme général - classe d'origine à statut élevé) et 61.7% ont des aspirations professionnelles de même niveau que leur classe sociale d'appartenance.

Tableau 9

Normalisation des tables de fréquences CP, CA et PA

*A. Les fréquences*

| | | CP | | | CA | | | PA | |
|---|---|---|---|---|---|---|---|---|---|
| | | P | | | A | | | A | |
| | | 1 | 2 | | 1 | 2 | | 1 | 2 |
| C | 1 | 841 | 49 | C | 1 676 | 214 | P 1 | 1335 | 646 |
| | 2 | 1140 | 188 | | 2 729 | 599 | 2 | 70 | 167 |

*B. Les probabilités*

| | | CP | | | CA | | | PA | |
|---|---|---|---|---|---|---|---|---|---|
| | | P | | | A | | | A | |
| | | 1 | 2 | | 1 | 2 | | 1 | 2 |
| C | 1 | .314 | .186 | C | 1 .309 | .191 | P 1 | .345 | .155 |
| | 2 | .186 | .314 | | 2 .191 | .309 | 2 | .155 | .345 |

On peut lire les tables du tableau 9 d'une autre façon. Les étudiants provenant de la classe sociale supérieure ont .314/.186 soit 1.917 chances contre 1 ou environ 2 chances contre 1 d'être inscrit au programme général plutôt qu'au programme professionnel, tandis que les étudiants provenant de la classe sociale inférieure n'ont qu'environ 1 chance contre 2 de faire de même. $C_1$ possède 1.61 chances contre 1 d'être dans $A_1$, et $C_2$ .62 chance contre 1; $P_1$ a 2.25 contre 1 d'être dans $A_1$ et $P_2$ .45 chance contre 1. En somme la classe sociale est un fort déterminant des chances pour un individu d'être inscrit au programme général plutôt qu'au programme professionnel, tandis que la fréquentation du programme général joue fortement sur l'expression d'aspirations professionnelles élevées. Ce qui nous amène à l'analyse de régression où, contrairement à la méthode employée jusqu'ici, une ou des variables sont considérées comme dépendantes.

Auparavant nous poursuivrons nos commentaires sur l'interprétation des coefficients $\lambda$. La notion de chances individuelles est simple d'interprétation et possède un lien direct avec les coefficients $\lambda$. Puisque $\lambda$ est le logarithme de $\tau$, nous nous servirons de $\lambda$ pour faciliter l'exposé. Par définition de $\lambda$ et en vertu de $P_{11} = P_{22}$ et $P_{12} = P_{21}$ dans les tables de fréquences normalisées du tableau 9:

$$(18) \qquad \tau_{11} = \left[\frac{(P_{11})^2}{(P_{12})^2}\right]^{\frac{1}{4}}$$

$$\tau_{11} = \left[\frac{P_{11}}{P_{12}}\right]^{\frac{1}{2}}$$

d'où les $\tau$ et les $\lambda$ sont des fonctions de proportions pondérées qui mesurent l'inégalité des chances d'appartenir à une catégorie de variables (conjointes) plutôt qu'à une autre.

Il est possible de concevoir les $\tau$ et les $\lambda$ comme des moyennes pondérées des chances d'être dans la case 1,1 plutôt que dans la case 1,2 et d'être dans la case 2,2 plutôt que dans la case 2,1 d'un tableau de contingence dichotomique à double entrée (ce raisonnement pourrait être étendu au cas général à N variables polytomiques). Puisque $\frac{p^{CA}_{11}}{p^{CA}_{12}}$, où P est une proportion et CA une variable conjointe, peut se définir comme les chances pour un individu d'être en $A_1$ plutôt qu'en $A_2$ si C=1 et $\frac{p^{CA}_{22}}{p^{CA}_{21}}$ comme les chances d'être en $A_2$ plutôt qu'en $A_1$ si C=2, alors $\tau^{CA}_{11}$ est la moyenne géométrique de deux fonctions de probabilités, c'est-à-dire $\frac{p^{CA}_{11}}{p^{CA}_{12}}$ et $\frac{p^{CA}_{22}}{p^{CA}_{21}}$. L'interprétation de cette moyenne est simple: il s'agit d'une fonction de la probabilité d'être dans la même classe en C et en A. Soit ici pour C et A:

$$(19) \quad \frac{p^{CA}_{11}}{p^{CA}_{12}} = \frac{.617}{.383} = 1.61$$

$$\frac{p^{CA}_{22}}{p^{CA}_{21}} = \frac{.617}{.383} = 1.61$$

$$\tau_{11} = (1.61 \times 1.61)^{\frac{1}{4}} \cdot 1.26 = 1.23^{1}$$

1. 1.2394 : valeur de $\lambda^{CA}$ du tableau 8.

La moyenne géométrique est la moyenne d'un produit. Puisque le produit croisé est un produit des chances d'être dans l'une ou l'autre des catégories (conjointes) de variables (conjointes), τ n'est rien d'autre que la moyenne de ce produit.

D) *Un analogue de l'analyse de régression*

Il est possible de mener l'équivalent d'une analyse de régression à l'aide de la même méthode, moyennant certaines transformations incluses dans l'ensemble ANOMHI. La procédure consiste essentiellement à expliquer le rapport des probabilités constitué par les chances qu'ont les individus appartenant à certaines catégories des variables indépendantes (ou explicatives) de se trouver dans une des catégories de la ou des variables dépendantes (ou à expliquer). L'intérêt du chercheur se portant sur les rapports entre les variables indépendantes et la ou les variables dépendantes, les paramètres d'interaction des premières seront maintenus constants. Ainsi dans le cas où la variable A dépend de C, S et P, les paramètres $\lambda^{C}_{i}$, $\lambda^{S}_{j}$, $\lambda^{P}_{k}$, $\lambda^{CS}_{ij}$, $\lambda^{CP}_{ik}$, $\lambda^{SP}_{jk}$ et $\lambda^{CSP}_{ijk}$ seront maintenus constants. Si dans l'équation 16 ils sont maintenus constants, nous obtenons l'équation de régression non saturée suivante:

$$(20) \qquad \psi^{CSP}_{ijk} = \Lambda^{\bar{A}}_{1} + \Lambda^{C\bar{A}}_{i1} + \Lambda^{S\bar{A}}_{j1} + \Lambda^{P\bar{A}}_{k1} + \Lambda^{CS\bar{A}}_{ij1}$$

où $\psi^{CSP}_{ijk}$ est le logarithme du rapport de probabilité à expliquer et où les $\Lambda$ jouent le même rôle que les $\lambda$ de l'équation 16. Dans le cas dichotomique les $\Lambda$ sont exactement égaux aux $\lambda$ de l'analyse générale. Evidemment, une équation saturée correspond à l'équation 15:

$$(21) \qquad \psi_{ijk}^{CSP} = A_{k}^{\overline{A}} + A_{ik}^{C\overline{A}} + A_{jl}^{S\overline{A}} + A_{kl}^{P\overline{A}} + A_{ijl}^{CS\overline{A}} + A_{ikl}^{CP\overline{A}} + A_{jkl}^{SP\overline{A}} + A_{ijkl}^{CSP\overline{A}}$$

On peut saisir de façon intuitive la signification des A lorsque l'on se convainc que l'analyse de régression proposée ici cherche à expliquer un rapport de proportion: si la variable dépendante A est dichotomique, ce que l'on explique est le rapport $A_1/A_2$ ou bien les chances d'appartenir à $A_1$ plutôt qu'à $A_2$. Puisque par exemple $\lambda_{ij}^{CA}$ (quand i=1 et j=1,2) sont des logarithmes des moyennes géométriques des chances d'appartenir à $A_1$ quand C=1 où à $A_2$ quand C=1, alors on comprend que:

$$(22) \quad A_{12}^{C\overline{A}} = \frac{\lambda_{11}^{CA} - \lambda_{12}^{CA}}{2} \quad (1)$$

Ce qui s'exprime simplement: connaissant C, le rapport $A_1/A_2$ où $\log A_1 - \log A_2$ pondéré par le nombre de termes dans l'équation, d'où $\frac{\log A_1 - \log A_2}{2}$ peut se mesurer par l'équation 22 en utilisant la logique générale qui nous permet d'utiliser $\lambda_{11}^{CA}$.

Les A du modèle saturé du tableau 10 sont équivalents aux $\lambda$ du modèle saturé évalués au tableau 8, puisque la variable dépendante A est dichotomique. Puisque $\lambda_{12}^{CA} = \frac{\lambda_{11}^{CA} - \lambda_{12}^{CA}}{2}$, or en vertu de l'égalité $\lambda_{11}^{CA} = -\lambda_{12}^{CA}$, alors $A_{12}^{CA} = \frac{\lambda_{11}^{CA} - (-\lambda_{11}^{CA})}{2} = \frac{2\lambda_{11}^{CA}}{2} = \lambda_{11}^{CA}$. En conséquence, nous ne nous étendrons pas outre-mesure sur ce que nous révèle les coefficients A.

(1) Ce qui est l'équivalent du logit de A (L.A. Goodman, 1970).

Le modèle 1 du tableau 11 peut s'interpréter comme un modèle de régression des variables C, S et P sur $\bar{A}$. La question alors posée par le chercheur est la suivante: quelle est la partie des chances pour un individu d'être situé en $A_1$ plutôt qu'en $A_2$ expliquée par C et SP. Il suppose que la classe sociale a un effet direct sur les aspirations professionnelles tandis que le sexe et le programme d'études forme une variable conjointe influençant le type d'aspirations professionnelles.

Cependant d'autres modèles plus économiques peuvent être construits, tout comme dans le cas de l'analyse générale précédente.

Tableau 10

$\Lambda$ et $\Lambda$ standardisé du modèle saturé $CSP\bar{A}$

| Paramètres relationnels | Tau | Lambda | Lambda standardisé |
|---|---|---|---|
| $A_2$ | 1.0434 | 0.2424 | .9721 |
| $C_1A_2$ | 1.2672 | 0.2368 | 5.4249 |
| $S_1A_2$ | 0.9496 | -0.0517 | -1.1844 |
| $P_1A_2$ | 1.4450 | 0.3681 | 8.4312 |
| $C_1S_1A_2$ | 1.0011 | 0.0011 | 0.0256 |
| $C_1P_1A_2$ | 0.9756 | -0.0247 | -0.5657 |
| $S_1P_1A_2$ | 1.0686 | 0.0664 | 1.5207 |
| $C_1S_1P_1A_2$ | 0.9872 | -0.0129 | -0.2944 |

Tableau 11

Modèles de régression vérifiés, leurs degrés de liberté et leurs Khi deux du rapport de vraisemblance

| | Modèles | Degrés de liberté | Khi deux du rapport de vraisemblance |
|---|---|---|---|
| (1) | CSP , $C\bar{A}$, $SP\bar{A}$ | 3 | .53 |
| (2) | CSP , $C\bar{A}$, $S\bar{A}$, $P\bar{A}$ | 4 | 3.60 |
| (3) | CSP , $C\bar{A}$, $P\bar{A}$ | 5 | 3.64 |
| (4) | CSP , $C\bar{A}$ | 6 | 104.60 |
| (5) | CSP , $P\bar{A}$ | 6 | 83.70 |
| (6) | CSP , $\bar{A}$ | | |
| (7) | C, S, P, $\bar{A}$ | 11 | 265.44 |
| (8) | Hypothèse d'équiprobabilité | 15 | 2079.92 |
| (9) | CSA, CPA, SPA | 2 | 3.39 |

$M_3$ du tableau 11 apparaît d'emblée comme le modèle le plus satisfaisant et le plus économique. Les paramètres relationnels de régression de $M_3$ se retrouvent dans le modèle théorique généré par l'analyse précédente. La significativité de l'apport de chacun des paramètres peut être mesurée par l'intermédiaire de l'analyse des $G^2$ pris comme estimés de la variation non expliquée par un modèle. Le tableau 12 donne le résultat de cette étude tandis que le tableau 13 converti ces résultats en coefficients de détermination multiple, multi-partielle ou partielle selon le cas.

Tableau 12

Analyse de la variation de $M_1$ et de $M_3$ du tableau 10

| | Modèles | Degrés de liberté | Khi deux du rapport de vraisemblance |
|---|---|---|---|
| (1) | Variation due à C, S, P, $\bar{A}$ non expliquée par $M_2$ | 4 | 3.60 |
| (1a) | Variation due à C, S, P, $\bar{A}$ non expliquée par $M_1$ | 3 | 0.53 |
| (1b) | Variation non expliquée par $\lambda_{SP\bar{A}}$ dans $M_1$ | 1 | 3.07* |
| (1) | Variation due à C, S, P, $\bar{A}$ | | 265.44* |
| (1a) | Variation due à C, S, P, $\bar{A}$ non expliquée par $M_3$ | 5 | 3.64 |
| (1b) | Variation expliquée par $M_3$ | 6 | 253.80* |
| (1a.1) | Variation due à C, S, P, $\bar{A}$ non expliquée par $M_4$ | 6 | 104.60* |
| (1a.1a) | Variation expliquée par $\lambda_{P\bar{A}}$ dans $M_3$ | 1 | 100.96* |
| (1a.2) | Variation due à C, S, P, $\bar{A}$ non expliquée par $M_5$ | 6 | 83.70* |
| (1a.2a) | Variation expliquée par $\lambda_{C\bar{A}}$ dans $M_3$ | 1 | 80.06* |

* Significatif à .05.

Il est possible d'appliquer ici la procédure de normalisation de F. Mosteller (1968). Le procédé donnera ici le même résultat que précédemment, aussi terminerons nous cette analyse de régression en notant le grand parallélisme pratique de l'utilisation de l'analyse générale et de l'analyse de régression d'une table de fréquences. L'interprétation des résultats est cependant sensiblement différente. En résumé, il est possible d'écrire $M_3$ du tableau 11 sous la forme usuelle d'un graphe:

$$\Lambda_{11}^{C\bar{A}} = .2146$$

$$\Lambda_{11}^{P\bar{A}} = .3658$$

où les valeurs des $\Lambda$ sont estimées à l'aide des fréquences $M_3$.

Tableau 13

Coefficients de détermination multiple, multipartielle et partielle

| | Valeurs de $R^2$ |
|---|---|
| $R^2_{f.M_3}$ | .994 |
| $R^2_{f(M_3)\ .\ \Lambda_1^C\ \Lambda_1^P\ \Lambda_1^S\ \Lambda_1^{\bar{A}}}$ | .969 |
| $R^2_{f\Lambda_{P\bar{A}}.M_4}$ | .958 |
| $R^2_{f\Lambda_{S\bar{A}}.M_3}$ | -.236 |

Il faut noter que l'opération de régression, consistant à estimer un rapport de fréquences (ici $A_1/A_2$), définit constant tout ce qui intervient entre les variables indépendantes, soit CSP dans cet exemple-ci. Cette constante possède un avantage majeur dans l'analyse: si le modèle qui exclut cette constante (donc qui contredit cette condition de régression) ne peut être rejeté (c'est-à-dire évalue de façon satisfaisante les fréquences observées), les modèles qui considèrent d'autre(s) variable(s) comme dépendante(s) que la variable dépendante sous étude sont plausibles. Le modèle 9 du tableau 11 étant non rejeté à .05, d'autres variables peuvent être considérées comme dépendantes: soit le programme d'études par exemple. Ainsi la conclusion suivante est légitime: les aspirations professionnelles et l'inscription dans un programme d'études s'influencent réciproquement. Il devient tout aléatoire de ne retenir pour fins d'analyse qu'une direction univoque de l'effet de l'une sur l'autre comme certaines études l'ont laissé entrevoir (W.H. Sewell et al, 1969).

*E) L'étude de plusieurs échantillons*[1]

L'étude qui précède a utilisé un seul niveau d'éducation. La compréhension du réseau de variables qui influence les aspirations scolaires n'est peut-être pas complète lorsqu'est considéré un seul niveau scolaire, aussi crucial est-il dans l'histoire scolaire de chacun des étudiants. Au Québec, le secondaire V, programme général, donne accès au niveau collégial (CEGEP) qui offre des programmes terminaux de formation technique et des programmes généraux qui ou-

1. Voir L.A. Goodman (1973).

vrent les portes de l'université. Il est donc possible d'écrire que la sélection des étudiants à l'entrée de l'université et que leur évacuation du système scolaire vers le marché du travail s'effectuent en deux moments distincts représentés par les résultats de leur passage au secondaire et au cégep. Une étude complète du système de sélection et d'évacuation des étudiants requiert au moins une analyse des deux niveaux d'éducation.

En conséquence, deux niveaux d'éducation seront étudiés, soient Secondaire V (E1) et CEGEP II (E2) du secteur anglophone des cohortes de l'étude A.S.O.P.E. Deux types de renseignements sont nécessaires pour mener à bien l'analyse: 1o la comparaison entre E1 et E2 étant donné un ensemble d'interactions des variables C, S, P et A et 2o le comportement particulier de E1 sur un ensemble d'interactions des variables C, S, P et A et de E2 sur un ensemble d'interactions de variables C, S, P et A.

On peut définir pour les échantillons des paramètres relationnels $\lambda$ ou $\Lambda$ tels qu'en les équations 15, 16, 20 ou 21, où sera prise en considération l'appartenance à l'échantillon $e = 1, \ldots, E$ quand E = nombre d'échantillons, ainsi:

$$(23)\quad \ln \tilde{F}_{ijkl} = \theta_e + \lambda^{C}_{ie} + \lambda^{S}_{je} + \lambda^{P}_{ke} + \lambda^{A}_{le} + \lambda^{CS}_{ije} + \lambda^{CP}_{ike} + \lambda^{CA}_{ile} + \lambda^{SP}_{jke} + \lambda^{SA}_{jle} + \lambda^{PA}_{kle} + \lambda^{CSP}_{jke} + \lambda^{CSA}_{ijke} + \lambda^{CPA}_{ikle} + \lambda^{SPA}_{kjle} + \lambda^{CSPA}_{ijkle}$$

décrit l'équation saturée pertinente.

Correspondent à la comparaison entre échantillons des coefficients qui considèrent: 1o des paramètres qui mesurent les interactions entre C, S, P et A sur l'ensemble des échantillons, tel que par exemple:

$$(24) \quad \lambda_i^C = \sum_e \frac{\lambda_{ie}^C}{n}$$

D'où les $\lambda$ généraux (c'est-à-dire afférant à l'ensemble des échantillons) sont égaux à la moyenne des $\lambda_e$. Et, 2o des paramètres qui mesurent la différence entre les $\lambda$ généraux et chacun des $\lambda_e$. L'observation de la valeur des $\lambda$ standardisés permet de vérifier 1o si les ($\lambda$: moyennes) et 2o si les ($\lambda$: différences de moyennes) sont significatifs. La somme de ces $\lambda$ plus une constante correspond à une estimation des $f_{ijkle}$:

$$(25) \quad \ln f_{ijkle} = \theta_e + \lambda_i^C + \lambda_j^C + \lambda_k^P + \lambda_l^A + \lambda_{le}^{CE} + \lambda_{je}^{SE} + \lambda_{ke}^{PE} + \lambda_{le}^{AE} + \lambda_{ij}^{CS} + \lambda_{ije}^{CSE} + \ldots + \lambda_{ijk}^{CSP} + \lambda_{ijke}^{CSPE} + \ldots + \lambda_{ijkle}^{CSPAE}$$

Puisque nous pouvons aussi bien utiliser la comparaison d'échantillons dans le cas d'une analyse générale des tableaux de contingence que dans le cadre d'une analyse de régression, l'utilisation de l'une ou l'autre des méthodes ne produit pas de problèmes spécifiques d'utilisation. Nous procéderons ici à une analyse de régression.

L'analyse de régression fournira trois listes de coefficients: la première liste décrit les paramètres de l'équation 25 modifiée, c'est-à-dire les paramètres de moyenne et de différence de moyennes. (Dans le texte E sera ajouté aux coefficients A); la deuxième liste donne les A du premier groupe d'étudiants (dans le texte aux A sont ajoutés la lettre E et l'indice 1) et finalement la troisième liste inclut les A du second groupe d'étudiants (les A sont écrits dans le texte avec la lettre E et l'indice 2). Le tableau 14 donne les valeurs de chacun des lambda. (Notez que par exemple: $A_{11}^{C\bar{A}} = \frac{A_{111}^{C\bar{A}E} - A_{112}^{C\bar{A}E}}{2}$ soit: $.1566 = \frac{(.2368 - .1177)}{2}$). Les A standardisés des moyennes des échantillons significatifs pris un à un, donc > à ±1.96 sont: $A_{1}^{\bar{A}}$, $A_{11}^{C\bar{A}}$, $A_{11}^{S\bar{A}}$, $A_{11}^{P\bar{A}}$, et $A_{111}^{SP\bar{A}}$. En somme, le modèle [CSP], $C\bar{A}$, $SP\bar{A}$ pourrait être validé. Il n'a d'intérêt en ce qu'il réintroduit le sexe comme variable explicative dans un modèle général qui prend pour objet d'observation le système de sélection des étudiants en son entier, plutôt qu'en un point particulier. Existent-ils des différences significatives entre les échantillons d'étudiants des niveaux d'étude secondaire V et CEGEP II? La distribution des aspirations professionnelles elle-même est significativement différente d'un échantillon à l'autre (remarquez que $A_{11}^{\bar{A}E} = A_{11}^{\bar{A}E} - A_{1}^{\bar{A}}$ soit: $-.2144 = .0424 - .2568$, de même $A_{11}^{C\bar{A}E}$ standardisé vaut 2.2624). Cependant, $A_{111}^{SP\bar{A}E}$ n'est pas significatif, confirmant par là l'importance du sexe comme variable discriminante dans le système général de sélection et d'évacuation. Le modèle [CSP], $C\bar{A}$, $P\bar{A}$, relatif au groupe de secondaire V seulement, ne serait donc

Tableau 14

Coefficients τ, Λ et Λ standardisé d'une analyse de régression à plusieurs échantillons

| Paramètres relationnels | Tau | Lambda | Lambda standardisé |
|---|---|---|---|
| $A_2$ | 1.2928 | 0.2568 | 7.2427 |
| $C_1A_2$ | 1.1695 | 0.1566 | 4.4160 |
| $S_1A_2$ | 0.8943 | -0.1117 | -3.1502 |
| $P_1A_2$ | 1.3653 | 0.3114 | 8.7811 |
| $E_1A_2$ | 0.8070 | -0.2144 | -6.0460 |
| $C_1S_1A_2$ | 1.0006 | 0.0006 | 0.0177 |
| $C_1P_1A_2$ | 0.9782 | -0.0220 | -0.6216 |
| $C_1E_1A_2$ | 1.0835 | 0.0802 | 2.2624 |
| $S_1P_1A_2$ | 1.1200 | 0.1134 | 3.1964 |
| $S_1E_1A_2$ | 1.0618 | 0.0600 | 1.6921 |
| $P_1E_1A_2$ | 1.0583 | 0.0567 | 1.5982 |
| $C_1S_1P_1A_1$ | 1.0101 | 0.0101 | 0.2847 |
| $C_1S_1E_1A_2$ | 1.0005 | 0.0005 | -0.0748 |
| $C_1P_1E_1A_2$ | 0.9973 | -0.0027 | -0.0748 |
| $S_1P_1E_1A_2$ | 0.9541 | -0.0470 | -1.3244 |
| $C_1S_1P_1E_1A_2$ | 0.9773 | -0.0229 | -0.6471 |

Tableau 14 (suite)

*Echantillon no. 1*

| Paramètres relationnels | Tau | Lambda | Lambda standardisé |
|---|---|---|---|
| $A_2$ | 1.0434 | 0.0424 | 0.9721 |
| $C_1A_2$ | 1.2672 | 0.2368 | 5.4249 |
| $S_1A_2$ | 0.9496 | -0.0517 | -1.1844 |
| $P_1A_2$ | 1.4450 | 0.3681 | 8.4312 |
| $C_1S_1A_2$ | 1.0011 | 0.0011 | 0.0256 |
| $C_1P_1A_2$ | 0.9756 | -0.0247 | -0.5657 |
| $S_1P_1A_2$ | 1.0686 | 0.0664 | 1.5207 |
| $C_1S_1P_1A_2$ | 0.9872 | -0.0129 | -0.2944 |

*Echantillon no. 2*

| Paramètres relationnels | Tau | Lambda | Lambda standardisé |
|---|---|---|---|
| $A_2$ | 1.6020 | 0.4713 | 8.5201 |
| $C_1A_2$ | 1.0794 | 0.0764 | 1.3808 |
| $S_1A_2$ | 0.8422 | -0.1717 | -3.1046 |
| $P_1A_2$ | 1.2901 | 0.2547 | 4.6053 |
| $C_1S_1A_2$ | 1.0001 | 0.0001 | 0.0020 |
| $C_1P_1A_2$ | 0.9808 | -0.0194 | -0.3505 |
| $S_1P_1A_2$ | 1.1739 | 0.1603 | 2.8985 |
| $C_1S_1P_1A_2$ | 1.0336 | 0.0330 | 0.5974 |

pas satisfaisant. Notre hypothèse de départ corrigée se verrait confirmer[1].

Quelques différences entre les deux groupes ont déjà été signalées. Quelle est l'étendue de ces différences? L'hypothèse d'hétérogénéité des groupes est confirmée par le modèle 3 du tableau 5. Le modèle 1, que nous retenons, contient un terme incluant la lettre E. Ce qui indique une différence significative sur ce paramètre entre les deux groupes. D'ailleurs $\Lambda_{111}^{CAE} = .2368$, tandis que $\Lambda_{112}^{CAE} = .0764$. Ces deux valeurs montrent que l'influence de la classe sociale sur les aspirations scolaires est beaucoup plus forte en secondaire V qu'au CEGEP; en fait, à ce dernier niveau scolaire elle est non significative ($\Lambda_{112}^{CAE}$ standardisé = 1.3808).

Puisque ces similitudes ou différences entre échantillons sont multiples, de multiples modèles peuvent être vérifiés. Notez cependant que si le modèle d'homogénéité complète ($M_3$ du tableau 5) est accepté, il est inutile de poursuivre une étude de différences. Par conséquent, ce modèle devrait être vérifié dès le début. De même si $M_4$ du tableau 15 est accepté seules les distributions des échantillons ou des groupes sur les variables prises une à une sont différentes. Autrement dit les paramètres relationnels d'interaction d'ordre 1 et plus, qui habituellement sont du premier niveau d'analyse d'intérêt, ne montrent pas de différences de groupes à groupes. Ce modèle devrait donc, lui aussi, être vérifié automatiquement.

---

1. Le modèle 3, E, CSPA n'est pas significatif. Il indique qu'un modèle significatif exige l'inclusion d'au moins une relation entre les paramètres C, S, P, A et E.

Dernière remarque:

$M_1$ du tableau 15 peut s'écrire de la façon suivante:

$$(26)\quad \psi_{111}^{CSPE} = A_1^{\bar{A}} + A_{11}^{\bar{A}E} + A_{11}^{C\bar{A}} + A_{111}^{C\bar{A}E} + A_{11}^{S\bar{A}} + A_{11}^{P\bar{A}} + A_{111}^{SP\bar{A}}$$

Puisque:

$$(27)\quad A_1^{\bar{A}E} = A_{11}^{\bar{A}} - A_1^{\bar{A}}$$

$$A_{11}^{C\bar{A}E} = A_{111}^{C\bar{A}} + A_{11}^{C\bar{A}}$$

$$A_{11}^{S\bar{A}} = \frac{A_{111}^{P\bar{A}} + A_{112}^{P\bar{A}}}{2}$$

$$A_{111}^{P\bar{A}} = \frac{A_{1111}^{P\bar{A}} + A_{1112}^{P\bar{A}}}{2}$$

$$A_{111}^{SP\bar{A}} = \frac{A_{1111}^{SP\bar{A}} + A_{1112}^{SP\bar{A}}}{2}$$

Alors:

$$(28)\quad \psi_{111}^{CSPE} = A_1^{\bar{A}} + A_{11}^{\bar{A}} + A_1^{\bar{A}} + A_{11}^{C\bar{A}} + A_{111}^{C\bar{A}} - A_{11}^{C\bar{A}} + A_{111}^{S\bar{A}} + \frac{A_{1111}^{S\bar{A}} + A_{1112}^{S\bar{A}}}{2}$$

$$+ \frac{A_{1111}^{P\bar{A}} + A_{1112}^{P\bar{A}}}{2} + \frac{A_{1111}^{SP\bar{A}} + A_{1112}^{SP\bar{A}}}{2}$$

et:

$$(29)\quad \psi_{111}^{CSPE} = 2A_{11}^{\bar{A}} + 2A_{111}^{C\bar{A}} + A_{111}^{S\bar{A}} + A_{112}^{S\bar{A}} + A_{111}^{P\bar{A}} + A_{112}^{P\bar{A}}$$

$$+ A_{1111}^{SP\bar{A}} + A_{1112}^{SP\bar{A}}$$

Tableau 15

Modèles vérifiés dans le cas de l'analyse de régression à plusieurs échantillons

| Modèles | | Degrés de liberté | Khi deux du rapport de vraisemblance |
|---|---|---|---|
| $(M_1)$ | CSPE , $C\bar{A}E$, $SP\bar{A}$ | 10 | 9.53 |
| $(M_2)$ | CSPE , $C\bar{A}$, $\bar{A}E$, $SP\bar{A}$ | 11 | 14.95 |
| $(M_3)$ | E, $CSP\bar{A}$ | 15 | 282.34 |
| $(M_4)$ | CE, SE, PE, $\bar{A}E$, $CSP\bar{A}$ | 11 | 30.54 |
| $(M_5)$ | C, S, P, $\bar{A}$, E | 26 | 551.27 |
| $(M_6)$ | Hypothèse d'équiprobabilité | 31 | 3402.20 |

Cette dernière équation permet de mieux comprendre la procédure d'analyse employée dans le cas d'une étude à deux échantillons ou plus. Les $\Lambda$ communs aux deux échantillons apparaissent identifiés dans l'équation, tandis que les $\Lambda$ particuliers à chacun d'eux sont pondérés par un facteur équivalent au nombre de présences dans les équations afférentes à E échantillons.

Cette équation peut facilement se recomposer en partant d'équations par échantillons en respectant la significativité des $\Lambda$. La lecture des valeurs de $\Lambda$ standardisés de chacun des échantillons peut aider en cette matière. Puisque par exemple, $\Lambda_{111}^{C\bar{A}}$, $\Lambda_{112}^{CA}$ seul $\Lambda_{111}^{C\bar{A}}$ apparaîtra dans l'équation. Le choix de l'un ou l'autre terme dépend de la valeur de $\Lambda_{11e}^{CA}$ dans chacun des échantillons.

F) *La quasi-indépendance*

Jusqu'ici nous n'avons traité que de tableaux de contingence complets. En certaines occasions, il est possible d'obtenir des cellules vides, soit parce que l'échantillon n'a pas tiré d'événements de la sorte, soit parce qu'il n'existe pas de ces événements dans la population, soit parce que cette cellule générée par le croisement des catégories de variables ne peut exister dans l'univers. Un autre type de cellule vide peut être distingué: si à l'analyse du tableau, il appert qu'une cellule présente un intérêt spécifique, il peut être avantageux de l'isoler du reste des cellules du tableau en la considérant comme nulle pour les besoins de l'analyse.

Le concept de quasi-indépendance permet d'étudier les tableaux tronqués. Il a été développé par L.A. Goodman (1961, 1963a, 1965, 1968), qui l'a appliqué à l'étude des tableaux de mobilité.

Comme précédemment, nous négligerons d'exposer les fondements de la théorie de la quasi-indépendance. Qui plus est, nous nous limiterons à l'analyse des tableaux bi-dimensionnels. La complexité des hypothèses de quasi-indépendance croît avec le nombre de variables, aussi nous référons le lecteur intéressé au chapitre 5 du livre de Y.M.M. Bishop et al. (1975), pour de plus amples détails.

Enfin, le lecteur peut prendre note que l'analyse des données soumises à la quasi-indépendance, même si elle peut se traduire à une équation de la même forme que l'équation 15, ne tire pas grand profit de l'utilisation de cette procédure.

Nous traiterons trois exemples: 1o le cas d'un tableau de contingence triangulaire où les cellules vides représentent des événements impossibles, 2o le cas où certaines cellules sont des zéros structuraux, tandis que certaines autres sont des zéros d'échantillonnage et 3o le cas où des considérations analytiques conduisent à traiter certaines cellules de façon spécifique. Ces trois exemples sont l'objet de l'attention de Y.M.M. Bishop et al. (1975). Nous les avons repris en vertu de leur clarté.

a. Les événements impossibles

Le tableau 16 représente la distribution de 121 patients atteints de maladies cardio-vasculaires plus ou moins graves (gravité de A à E) à leur entrée à l'hôpital (variable I) et leur état (de A à E) à leur sortie de l'hôpital (variable E). Puisque aucun patient ne sort de l'hôpital si son état s'aggrave, à moins de mortalité, une partie des cellules du tableau ont des zéros structuraux.

Tableau 16

Etat à l'entrée et à la sortie de l'hôpital d'un échantillon de malades atteints de maladies cardio-vasculaires

| | Etat à la sortie (E) | | | | |
|---|---|---|---|---|---|
| Etat à l'entrée (I) | A | B | C | D | E |
| E | 11 | 23 | 12 | 15 | 8 |
| D | 9 | 16 | 4 | 1 | - |
| C | 6 | 8 | 4 | - | - |
| B | 4 | 5 | - | - | - |
| A | 5 | - | - | - | - |

Source: Y.M.M. Bishop et S.E. Fienberg (1969).

Les estimés et le Khi deux de vraisemblance maximum, avec 6 degrés de liberté (RC-(R-1) + (C-1)-1-Z) où R est le nombre de rangées C le nombre de colonnes et Z le nombre de zéros structuraux, sont donnés au tableau 17. Il n'y a donc pas quasi-indépendance. L'examen des déviations des estimés des observations montrerait que la cellule (1,1) contribue plus que sa part à ce $X^2$. On pourrait imaginer un autre modèle qui exclurait les cas (1,1) soit ceux qui entrent à l'hôpital dans un état avancé de morbidité (catégorie E), pour en sortir avec le moindre degré de morbidité (catégorie A). On pourrait, dès lors, émettre l'hypothèse d'un seuil de morbidité E qui ne permet pas de retour à un état de santé meilleur que B.

Tableau 17

Estimés des fréquences du tableau 16 soumis à la quasi-indépendance

| I | A | B | E C | D | E |
|---|---|---|---|---|---|
| E | 14.99 (-3.99) | 23.98 (-0.98) | 11.30 (0.70) | 10.72 (4.28) | 8.00 (0.00) |
| D | 7.37 (4.17) | 11.80 (6.68) | 5.56 (-1.56) | 5.28 (-4.23) | |
| C | 4.17 (1.83) | 6.68 (-2.68) | 3.15 (0.85) | | |
| B | 3.46 (0.54) | 5.54 (-0.54) | | | |
| A | 5.00 (0.00) | | | | |

Les déviations sont inscrites sous les estimés. ANOMHI donnera la valeur zéro aux cellules vides originales.

$X^2$ = 12.46; 6 degrés de liberté

b. Les zéros structureux et les zéros d'échantillonnage

La tribu des Purums du centre de l'Inde est organisée en clans exogamiques. Le tableau 18 donne le nombre de mariages entre chacun des clans. Les cellules (2,3) et (5,4) ont des zéros d'échantillonnage puisque des mariages entre membres de ces clans ont été observés dans d'autres échantillons. La cellule (5,5) représente un cas d'exception puisqu'il s'agit d'un mariage endogamique. Cependant, T. Das (1945) conclut de ce mariage qu'un clan est en formation parmi les Khekang.

Nous posons la question suivante: la distribution des mariages entre clans exogamiques se fait-elle au hasard? L'hypothèse de quasi-indépendance permet de répondre à cette question.

Tableau 18

Mariages entre clans exogamiques de la tribu des Purums et estimés des fréquences selon la quasi-indépendance d'après 5 itérations

| Clan de la femme | Clan du mari | | | | |
|---|---|---|---|---|---|
| | Marrim | Maran | Parpa | Thao | Khekang |
| Marrim | | 5 (10.79) | 17 (10.68) | | 6 (6.54) |
| Maran | 5 (4.86) | | 0 (6.57) | 16 (7.55) | 2 (4.02) |
| Parpa | | 2 (8.38) | | 10 (9.54) | 11 (5.08) |
| Thao | 10 (10.39) | | | | 9 (8.60) |
| Khekang | 6 (5.74) | 20 (7.83) | 8 (7.76) | 0 (8.23) | 1 (4.75) |

$\chi^2$ = 76.25; 8 degrés de liberté

Source: T. Das (1945).

Le $G^2$ avec 8 degrés de liberté ne nous permet pas d'accepter l'hypothèse de quasi-indépendance. En conséquence, des règles d'échange entre clans gouvernent probablement les mariages en sus des obligations exogamiques et des exclusions totales.

c. L'observation de cellules spécifiques

Les données que nous traiterons ici, ont été abondamment analysées entre autres par H.C. White (1963) et L.A. Goodman (1965, 1968). Elles proviennent d'une étude de D.V. Glass (1954) sur la mobilité sociale en Angleterre.

L'observation des estimés des fréquences du tableau 19 sous indépendance totale indique une forte déviation dans chacune des cellules de la diagonale. L'hypothèse de quasi-indépendance dans les cellules hors-diagonales pourrait être vérifiée. Pourrait être testée dans un même moment une hypothèse de quasi-indépendance qui impliquerait une distinction entre la mobilité ascendante et la mobilité descendante. Le tableau 20 montre que l'hypothèse uniforme de quasi-indépendance dans toutes les cellules hors-diagonales n'est pas acceptable. Le tableau 21 qui est le résultat de la combinaison de deux hypothèses de quasi-indépendance, une pour chacun des deux types de mobilité, démontre que la mobilité n'est pas un phénomène univoque puisque deux hypothèses doivent être posées: l'une pour la mobilité ascendante et l'autre pour la mobilité descendante.

Tableau 19

Mobilité sociale en Angleterre et estimés des fréquences sous hypothèse d'indépendance totale

| Statut du père | Statut du fils 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 50 (3.77) | 45 (17.89) | 8 (16.79) | 18 (52.27) | 8 (38.29) |
| 2 | 28 (14.46) | 174 (68.63) | 84 (64.42) | 154 (200.55) | 55 (146.94) |
| 3 | 11 (15.13) | 78 (71.82) | 110 (67.41) | 223 (209.87) | 96 (153.77) |
| 4 | 14 (44.10) | 150 (209.35) | 185 (196.51) | 714 (611.79) | 447 (448.25) |
| 5 | 0 (25.55) | 42 (121.31) | 72 (113.87) | 320 (345.52) | 411 (259.75) |

$x^2$ = 842.30, 16 degrés de liberté

Signification des catégories:

1 = professionnels et administrateurs
2 = cadres
3 = contremaîtres
4 = travailleurs qualifiés
5 = travailleurs non qualifiés

Si le $X^2$ du tableau 19 est considéré comme la variation maximum, la somme des $X^2$ du tableau 21 peut-être imputée à la diagonale et considérée comme la partie expliquée de la variation du $X^2$ d'indépendance totale. En conséquence, il est permis de conclure que l'héritage du statut social explique la distribution des fils dans différentes catégories du statut social puisqu'il y a manifestement dépendance du statut du fils sur le statut du père lorsque le premier en hérite.

Tableau 20

Quasi-indépendance uniforme hors la diagonale

| Statut du père | Statut du fils 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | 9.54 | 11.02 | 38.49 | 19.95 |
| 2 | 6.56 | | 49.93 | 173.23 | 90.32 |
| 3 | 8.52 | 56.06 | | 225.17 | 117.25 |
| 4 | 27.81 | 180.96 | 209.04 | | 378.48 |
| 5 | 10.40 | 68.44 | 79.06 | 275.11 | |

$\chi^2$ = 249.43; 11 degrés de liberté

Tableau 21

Quasi-indépendance spécifique pour chacun des types de mobilité

| Statut du père | Statut du fils 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | 45.00 | 9.57 | 17.43 | 7.02 |
| 2 | 27.99 | | 82.43 | 150.14 | 60.45 |
| 3 | 7.54 | 61.44 | | 227.43 | 91.58 |
| 4 | 13.16 | 142.11 | 193.69 | | 446.95 |
| 5 | 4.30 | 46.45 | 63.31 | 320.00 | |

$\chi^2$ de quasi-indépendance de mobilité ascendante: 12.60; 3 d.l.

$\chi^2$ de quasi-indépendance de mobilité descendante: 1.36; 3 d.l.

$\chi^2$ contribution de la diagonale: 842.30 - (12.60+1.36) = (11-(3+3)) d.l.

### a) L'indépendance régionale

L'hypothèse d'indépendance régionale permet de vérifier si des catégories de variables dans un tableau de contingence peuvent être amalgamées (F. Béland, 1978; F. Béland et J.J. Fortier, 1978). Essentiellement, l'indépendance régionale, lorsqu'elle ne peut être rejetée, permet de réduire le nombre de catégories d'une ou de plusieurs des variables étudiées. Cette réduction peut être désirable tant du point de vue théorique que du point de vue pratique dans la mesure où un tableau de variables de peu de catégories s'étudient plus facilement qu'un tableau de variables possédant un grand nombre de catégories.

Comme exemple, les données du tableau 22 seront retenues. Il est en effet plus simple de traiter du cas bivarié. L'extension au cas multivarié peut être considérée comme directe à toutes fins pratiques. Il en est longuement question dans un article de F. Béland et J.J. Fortier (1978). Le chercheur intéressé pourra consulter ce texte.

Amalgamer des catégories d'une ou de plusieurs variables d'un tableau de contingence implique que les interactions entre ces catégories sont de peu d'intérêt pour le chercheur même si l'association entre les variables veut être retenue. En ce sens, des régions sont dessinées dans un tableau, régions où l'indépendance entre les cellules d'une même région est posée comme hypothèse. L'indépendance ne concerne donc que des parties de tableau. En ce sens, elle est régionale. Il est important de noter que l'amalgamation s'appli-

que ici aux catégories. La conséquence de cette amalgamation est la création de régions dans le tableau. Ainsi le tableau 22 se voit imposer en pointillé une structure régionale dérivée de l'amalgamation des catégories 1 et 2 d'une part, 3,4 et 5 d'autre part de chacune des classifications d'occupation des pères et des fils. Si l'hypothèse d'indépendance régionale ne pouvait être rejetée, le tableau 22 pourrait donner lieu à un tableau dit régionalisé où seules les fréquences régionales peuvent être retenues. Puisque les fréquences régionales sont l'addition des fréquences de chacune des régions, au lieu d'un tableau bi-varié à 5x5 = 25 fréquences, un tableau dichotomique (2x2 = 4 fréquences) suffirait pour expliquer la structure de la mobilité sociale.

Tableau 22

Structure régionale imposée aux données du tableau

| Statut du père | Statut du fils 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 50 | 45 | 8 | 18 | 8 |
| 2 | 28 | 174 | 84 | 154 | 55 |
| 3 | 11 | 78 | 110 | 223 | 96 |
| 4 | 14 | 150 | 185 | 714 | 447 |
| 5 | 0 | 42 | 72 | 320 | 411 |

L'hypothèse d'indépendance régionale s'exprime aisément en fonction des modèles log-linéaires. L'équation saturée d'un tableau dichotomique tel le tableau 22 est:

$$(30) \quad \ln f_{ij} = \theta + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_{ij}^{(12)}$$

Les hypothèses d'indépendance régionale, telle qu'elles sont définies ci-haut, c'est-à-dire comme des hypothèses sur des amalgamations de catégories qui donnent lieu à des réductions de tableau, sont en fait des hypothèses d'égalité des paramètres d'interaction $\lambda_{ij}^{(12)}$ inclus dans chacune des régions. Ainsi au tableau 22 où les régions sont définies en pointillé, l'indépendance régionale implique:

$$(31\text{-}1) \quad \lambda_{11} = \lambda_{12} = \lambda_{21} = \lambda_{22}$$

$$(31\text{-}2) \quad \lambda_{13} = \lambda_{14} = \lambda_{23} = \lambda_{24} = \lambda_{33} = \lambda_{34}$$

$$(31\text{-}3) \quad \lambda_{31} = \lambda_{32} = \lambda_{41} = \lambda_{42} = \lambda_{51} = \lambda_{52}$$

$$(31\text{-}4) \quad \lambda_{43} = \lambda_{44} = \lambda_{53} = \lambda_{54} = \lambda_{63} = \lambda_{64} = \lambda_{73} = \lambda_{74} = \lambda_{83} = \lambda_{84}$$

Il est clair qu'il n'y a finalement que 4 valeurs originales de $\lambda_{ij}$ dans ce cas. Mais puisque selon cette hypothèse, le tableau 22 se réduit à un tableau 2x2, un seul paramètre $\lambda_{ij}$ est nécessaire.

Si le paramètre d'association appartenant au tableau réduit s'écrit $\lambda_{rc}$ où r = 1,2 et c = 1,2, l'équation log-linéaire de l'hypothèse d'indépendance régionale s'écrit:

$$(32) \quad \ln f_{ij} = 0 + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_{rc}^{(12)}$$

Les hypothèses d'indépendance régionale peuvent toujours s'écrire de cette façon. L'ensemble des paramètres non affectés par l'amalgamation possède autant de termes que prévu dans le tableau original, tandis que l'ensemble des termes appartenant au tableau réduit possède autant de catégories que prévu dans ce tableau. Ici le paramètre $\lambda_{rc}^{(12)}$ ne possède qu'un seul terme indépendant soit $\lambda_{11}^{(12)}$ puisque lui seul est nécessaire dans un tableau 2x2, conséquence de la définition même des paramètres dans les équations log-linéaires qui veut que:

$$(33) \quad \sum_i \lambda_{ij}^{(12)} = \sum_j \lambda_{ij}^{(12)} = 0$$

quand (i,j) ou (r,c) sont des indices appropriés.

Les degrés de liberté se calculent donc aisément dans le cas de l'indépendance régionale. Il s'agit toujours du nombre total de paramètres dans l'équation saturée, soient 24 dans le cas du modèle moins le nombre de paramètres non redondants, soient 9, pour un total de 15 degrés de liberté. La même procédure s'applique dans le cas multivarié. Dans le cas qui nous préoccupe, les paramètres des marginales du tableau ne sont pas affectées par l'hypothèse d'indépendance régionale, soit un total de 8 paramètres plus 1 paramètre $\lambda_{ij}$ nécessaire pour tenir compte de l'association entre l'occupation des pères et celle des fils. Le total est donc de 9 paramètres non-redondants.

Le tableau 24 donne les valeurs des $X^2$ et leurs degrés de liberté pour cinq hypothèses de regroupements. Aucune d'entre elles ne peut être retenue. Cependant, le coefficient d'efficacité de l'hypothèse de réduction construit de façon très analogue au coefficient de détermination partielle, indique que la réduction R2 à deux catégories est presque aussi efficace que la réduction R3 à trois catégories.

Le tableau 25 est une illustration des égalités 31. Il est clair que les $\lambda_{ij}$ d'une même région sont égaux entre eux. Pour une étude des rapports entre les $\lambda_{ij}$ de l'hypothèse d'indépendance régionale et ceux du tableau réduit correspondant, une lecture des textes cités est recommandée. Les rapports entre ces deux types de paramètres ne sont pas immédiats et dépendent d'une pondération par le nombre de catégories et du tableau original et du tableau réduit. Il y aura égalité entre eux lorsque le rapport entre les deux nombres est constant pour chacune des variables.

Tableau 24

$G^2$ et leurs degrés de liberté de 4 regroupements des catégories des données du tableau.

| Regroupements* | $G^2$ | d.l. | Coefficients d'efficacité |
|---|---|---|---|
| R1 [1],[2 3 4 5] | 629.90 | 15 | 0.252 |
| R2 [1 2], [3 4 5] | 422.27 | 15 | 0.499 |
| R3 [1 2], [3], [4 5] | 304.98 | 12 | 0.510 |
| R4 [1], [2 3], [4 5] | 290.67 | 12 | 0.523 |

* Les regroupements des catégories d'occupations des pères sont identiques à ceux des fils.

Tableau 25

Coefficients $\lambda_{ij}$ du regroupement R2

| | | Statut du fils | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Statut du père | 1 | .750 | .750 | -.500 | -.500 | -.500 |
| | 2 | .750 | .750 | -.500 | -.500 | -.500 |
| | 3 | -.500 | -.500 | .333 | .333 | .333 |
| | 4 | -.500 | -.500 | .333 | .333 | .333 |
| | 5 | -.500 | -.500 | .333 | .333 | .333 |

## II. ANOMHI: Un programme en langage APL.SVA pour l'étude des tableaux de contingence multivariés

Le programme ANOMHI (Analyse nominale hiérarchique de fréquences) permet une analyse détaillée de tout tableau de contingence. Essentiellement, il met à la disposition de l'utilisateur les méthodes d'analyse log-linéaire des tableaux de contingence développées entre autres par L.A. Goodman (1968, 1971, 1972a, 1972b, 1973).

Ecrit en langage APL.SVA, ANOMHI n'utilise que deux des fonctions propres au système APL du Centre de traitement de l'information de l'Université Laval. ANOMHI est entièrement conversationnel et ne requiert aucune connaissance préalable du langage APL.SVA. Certaines notions de base peuvent cependant être utiles si des diff'cul-

tés causées par des déficiences du système, ou certaines erreurs commises par l'utilisateur, se présentent[1]. Certaines de ces erreurs seront détectées automatiquement par ANOMHI. L'utilisateur sera appelé à corriger l'erreur selon une procédure simple à suivre. En d'autres circonstances il sera appelé à vérifier ses données ou instructions, puis à les corriger s'il y a lieu.

Pour utiliser ANOMHI sans plus de détails, il suffit de copier les sous-routines qui sont en appendices, dans un espace de travail, d'appeler le bloc de travail puis le programme ANOMHI si l'utilisateur n'a pas pris soin d'exécuter un [illegible] lors de la procédure de copie du programme.

A) *L'entrée des données*

Le programme ANOMHI permet de traiter des tableaux de contingence à N dimensions (N>1) dont les variables ont C catégories (de façon générale: C>1). Le tableau devra toujours avoir la forme du tableau 25. ANOMHI ne peut traiter que des tableaux de cette forme.

Les fréquences seront entrées les unes à la suite des autres, chacune séparée d'un espace vide dans l'ordre indiqué par la flèche au tableau 25.

---

1. L'index du livre de L. Gilman et A.J. Rose (1974) permet de référer rapidement l'utilisateur à la solution de la difficulté sans pour autant qu'il soit familier avec le langage APL.

Tableau 25

Disposition des fréquences pour leur entrée dans le programme ANOMHI. Cas d'un tableau à quatre variables.[1]

| VARIABLES | | | D | | | |
|---|---|---|---|---|---|---|
| A | B | C | 1 | 2 | 3 | 4 |
| 1 | 1 | 1 | 1111 → | 1112 → | 1113 → | 1114 |
| 1 | 1 | 2 | → 1121 → | 1122 → | 1123 → | 1124 |
| 1 | 1 | 3 | → 1131 → | 1132 → | 1133 → | 1134 |
| 1 | 2 | 1 | → 1211 → | 1212 → | 1213 → | 1214 |
| 1 | 2 | 2 | → 1221 → | 1222 → | 1223 → | 1224 |
| 1 | 2 | 3 | → 1231 → | 1232 → | 1233 → | 1234 |
| 2 | 1 | 1 | → 2111 → | 2112 → | 2113 → | 2114 |
| 2 | 1 | 2 | → 2121 → | 2122 → | 2123 → | 2124 |
| 2 | 1 | 3 | → 2131 → | 2132 → | 2133 → | 2134 |
| 2 | 2 | 1 | → 2211 → | 2212 → | 2213 → | 2214 |
| 2 | 2 | 2 | → 2221 → | 2222 → | 2223 → | 2224 |
| 2 | 2 | 3 | → 2231 → | 2232 → | 2233 → | 2234 |

1. Les variables A et B sont dichotomiques, la variable C est trichotomique et la variable D possède quatre catégories. Il eut été possible d'inscrire les variables sous un autre ordre, par exemple BDCA, DCAB, etc. en prenant soin d'organiser la nomenclature des catégories en fonction de l'ordre des variables de telle sorte que les catégories de l'avant dernière variable varient en fonction de la précédente, et ainsi de suite comme il est illustré dans ce tableau.

Mais avant d'entrer les fréquences un certain nombre de renseignements seront demandés. Le tableau de contingence pourra avoir été conservé dans un bloc de travail. Une question sera posée à cet effet. Il suffira de répondre par un *OUI* ou par un *NON*. Le tableau de contingence aura été entré sous la forme d'un vecteur (suite de chiffres) dans l'ordre indiqué au tableau 25. Si le tableau est conservé dans un bloc de travail, l'utilisateur devra copier son vecteur de fréquences (ou vecteur de données) dans le programme ANOMHI en inscrivant *PCOPY* suivi, entre guillemets, du nom du bloc de travail où est conservé le vecteur et du nom de la variable sous laquelle se trouve les fréquences. Si le bloc se nomme *DONNEES* et la variable *FR*, l'utilisateur inscrira:

*PCOPY'DONNEES FR'*

Une seconde question lui demandera d'inscrire les données. Il devra alors écrire *FR*. Le nom de variable *MD* est réservé à ANOMHI. Ne jamais conservé un vecteur intitulé *MD*.

L'utilisateur qui désire inscrire par la procédure d'ANOMHI son vecteur de données sera habilité à utiliser le mode d'entrée prévu à cette fin. Après la question *INSCRIVEZ VOTRE VECTEUR DE DONNEES*: un signe □: apparaîtra. En actionnant la touche RETURN, le même signe réapparaîtra. La fin de la procédure d'entrée du vecteur sera signalée par ⍳0 (vecteur vide).

En réponse à la question pertinente le nombre de catégories de chacune des variables est entré selon l'ordre des variables du ta-

bleau de contingence. Au tableau 25 la variable *A* est la première variable, la variable *B* la seconde, .... la variable *D* la dernière.

Une erreur pourra s'être glissée jusqu'ici puisqu'à un nombre de catégories doit correspondre un nombre fixe de fréquences. Toute erreur de ce genre sera signalée et l'utilisateur devra recommencer le tout.

Les variables seront représentées par des caractères que l'utilisateur choisira. Il est recommandé d'utiliser les caractères de l'alphabet latin[1]. Aucun espace ne devra apparaître entre les caractères. Un seul caractère doit être utilisé par variable.

*EXEMPLE* 1.1: *ENTREE DU VECTEUR DE DONNEES*

```
      )LOAD ANOMNI
EN DATE DU 10.17.31 02/24/78
ANALYSE NOMINALE HIERARCHIQUE DE FREQUENCES
PROGRAMME ECRIT PAR:                          SESSION DU 9 12 78
FRANCOIS BELAND                               DEBUTE A: 12:29:43 PM
A.S.O.P.E. 1978
VOULEZ-VOUS COPIER VOTRE VECTEUR D'UN BLOC DE TRAVAIL?
NON
 INSCRIVEZ VOTRE VECTEUR DE DONNEES:
[]:
     1 2 3 4
[]:
     5 6 7 8
[]:
     9 10 11 12
[]:
     13 14 15 16
```

1. Si des hypothèses d'indépendance régionale sont vérifiées, exclure les lettres I, O, E et H des caractères disponibles.

```
□:
      17 18 19 20
□:
      21 22 23 24
□:
      25 26 27 28
□:
      29 30 31 32
□:
      33 34 35 36
□:
      37 38 39 40
□:
      41 42 43 44
□:
      45 46 47 48
□:
```

*INSCRIVEZ LE NOMBRE DE CATEGORIES DE CHACUNE DE VOS VARIABLES:*
□: 2 2 3 4
*INSCRIVEZ LES CARACTERES QUI REPRESENTENT VOS DONNEES:*
*ABCD*

*LES FREQUENCES*

| | | | |
|---|---|---|---|
| 1.00 | 2.00 | 3.00 | 4.00 |
| 5.00 | 6.00 | 7.00 | 8.00 |
| 9.00 | 10.00 | 11.00 | 12.00 |
| 13.00 | 14.00 | 15.00 | 16.00 |
| 17.00 | 18.00 | 19.00 | 20.00 |
| 21.00 | 22.00 | 23.00 | 24.00 |
| 25.00 | 26.00 | 27.00 | 28.00 |
| 29.00 | 30.00 | 31.00 | 32.00 |
| 33.00 | 34.00 | 35.00 | 36.00 |
| 37.00 | 38.00 | 39.00 | 40.00 |
| 41.00 | 42.00 | 43.00 | 44.00 |
| 45.00 | 46.00 | 47.00 | 48.00 |

Enfin, une procédure de correction des données est disponible. La procédure de correction d'une ou de plusieurs fréquences est simple à suivre: il s'agit de répondre par *OUI* ou par *NON*, selon le cas, à la question qui sera posée. Il suffit par la suite d'entrer la valeur erronnée de la fréquence puis, après la question appropriée, d'inscrire la valeur juste. Plusieurs fréquences peuvent être corri-

gées à la fois. Il faut cependant noter que si deux ou plusieurs fréquences dont une seule est à corriger prend la même valeur, la procédure employée considère que les unes et les autres sont sujettes à correction. Il suffira de corriger la première fréquence dans l'ordre de son entrée dans le vecteur de données en premier lieu, puis la seconde, et ainsi de suite. De façon générale, l'ordre d'entrée des valeurs corrigées doit être le même que l'ordre d'entrée des valeurs erronnées.

*EXEMPLE 1.2: CORRECTION DE FREQUENCES*

```
VERIFIEZ LES FREQUENCES, AVEC-VOUS COMMIS DES ERREURS?
OUI
QUELLE(S) EST (SONT) LA (LES) VALEUR(S) DE LA (DES) FREQUENCE(S) A
CORRIGER?
□:
     14 18
PAR QUELLE(S) VALEUR(S) EST-ELLE (SONT-ELLES) REMPLACEE(S)?
□:
     8 25
LES FREQUENCES:

     1.00     2.00     3.00     4.00
     5.00     6.00     7.00     8.00
     9.00    10.00    11.00    12.00
    13.00     8.00    15.00    16.00
    17.00    25.00    19.00    20.00
    21.00    22.00    23.00    24.00
    25.00    26.00    27.00    28.00
    29.00    30.00    31.00    32.00
    33.00    34.00    35.00    36.00
    37.00    38.00    39.00    40.00
    41.00    42.00    43.00    44.00
    45.00    46.00    47.00    48.00

VERIFIEZ LES FREQUENCES, AVEC-VOUS COMMIS DES ERREURS?
NON
```

Si au moins une des fréquences est plus petite ou égale à 5, il sera possible d'obtenir un estimé pseudo-bayséen (S.E. Fienberg et P.W. Holland, 1973; K.W. Smith, 1976) des fréquences soit sous l'hypothèse d'un modèle log-linéaire à priori, soit à l'aide d'un second échantillon de la même ou d'une autre population distribué selon les mêmes variables et les mêmes catégories que celles du tableau des fréquences originales. Des questions spécifiques seront posées, puis les fréquences corrigées seront imprimées. Enfin l'utilisateur aura le choix entre ses données originales et les fréquences corrigées pour la suite des travaux. Le mode d'emploi de la procédure d'estimation des fréquences par un modèle log-linéaire est ici le même que celui exposé plus loin. Enfin, la statistique K est un indicateur de la distance entre les fréquences originales et les fréquences pseudo-bayséennes. K tend vers l'infini quand le rapport est parfait entre les fréquences originales et les fréquences pseudo-bayséennes (Y.M.M. Bishop et al., 1975).

*EXEMPLE 1.3:* *ESTIMES BAYSEENS LES PETITES FREQUENCES PAR UN MODELE LOG-LINEAIRE A PRIORI*

```
UNE OU DES FREQUENCES SONT PETITES. DESIREZ-VOUS UNE CORRECTION BAYSEENNE?
OUI
LES PROBABILITES A PRIORI SONT-ELLES DETERMINEES PAR UN MODELE LOG-LINEAIRE?
OUI
***ATTENTION: N'ENTREZ QU'UN SEUL MODELE***
VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) DE QUASI-INDEPENDANCE?
NON
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
[]:
     .01
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
```

(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
ABC, ABD, ACD, BCD
VERIFIEZ VOS ENTREES. AVEZ-VOUS COMMIS DES ERREURS?
NON

LES FREQUENCES CORRIGEES:

| | | | |
|---|---|---|---|
| 1.79 | 1.51 | 3.02 | 3.69 |
| 4.69 | 6.80 | 6.74 | 7.77 |
| 8.52 | 9.69 | 11.24 | 12.54 |
| 12.21 | 8.49 | 14.98 | 16.31 |
| 17.31 | 24.20 | 19.26 | 20.23 |
| 21.48 | 22.31 | 22.76 | 23.46 |
| 24.21 | 26.49 | 26.98 | 28.31 |
| 29.31 | 29.20 | 31.26 | 32.23 |
| 33.48 | 34.31 | 34.76 | 35.46 |
| 37.79 | 37.51 | 39.02 | 39.69 |
| 40.69 | 42.80 | 42.74 | 43.77 |
| 44.52 | 45.69 | 47.24 | 48.54 |

CONSTANTE K: 13811511.069

DESIREZ-VOUS UTILISER LES DONNEES ORIGINALES?
OUI

FIN DE LA PROCEDURE D'ENTREE DU VECTEUR DE DONNEES:

EXEMPLE 1.4: <u>ESTIMES BAYSEENS DES PETITES FREQUENCES AVEC UN AUTRE ECHANTILLON COMME ESTIME PSEUDO-BAYSEEN</u>

UNE OU DES FREQUENCES SONT PETITES. DESIREZ-VOUS UNE CORRECTION BAYSEENNE?
OUI
LES PROBABILITES A PRIORI SONT-ELLES DETERMINEES PAR UN MODELE LOG-LINEAIRE?
NON
ENTREZ LES FREQUENCES D'UN AUTRE ECHANTILLON:
□:
5 6 7 8
□:
9 10 11 12
□:
13 14 15 16
□:
17 18 19 20
□:
21 22 23 24
□:
25 26 27 28
□:
29 30 31 32
□:
33 34 35 36

```
[]:
    37 38 39 40
[]:
    41 42 43 44
[]:
    45 46 47 48
[]:
    49 50 61 62
[]:
    10
```

*LES FREQUENCES:*

| | | | |
|---|---|---|---|
| 5.00 | 6.00 | 7.00 | 8.00 |
| 9.00 | 10.00 | 11.00 | 12.00 |
| 13.00 | 14.00 | 15.00 | 16.00 |
| 17.00 | 18.00 | 19.00 | 20.00 |
| 21.00 | 22.00 | 23.00 | 24.00 |
| 25.00 | 26.00 | 27.00 | 28.00 |
| 29.00 | 30.00 | 31.00 | 32.00 |
| 33.00 | 34.00 | 35.00 | 36.00 |
| 37.00 | 38.00 | 39.00 | 40.00 |
| 41.00 | 42.00 | 43.00 | 44.00 |
| 45.00 | 46.00 | 47.00 | 48.00 |
| 49.00 | 50.00 | 51.00 | 52.00 |

*VERIFIEZ LES FREQUENCES. AVEZ-VOUS COMMIS DES ERREURS?*
*NON*
*LES FREQUENCES CORRIGEES:*

| | | | |
|---|---|---|---|
| 3.37 | 4.09 | 4.79 | 5.51 |
| 6.50 | 7.31 | 8.07 | 8.87 |
| 9.47 | 10.21 | 11.09 | 11.91 |
| 12.52 | 13.41 | 14.21 | 15.06 |
| 16.59 | 17.59 | 18.46 | 19.39 |
| 19.91 | 20.94 | 21.83 | 22.79 |
| 23.17 | 24.25 | 25.15 | 26.14 |
| 18.18 | 19.48 | 30.46 | 31.55 |
| 31.95 | 33.19 | 34.10 | 35.32 |
| 35.81 | 37.10 | 38.14 | 39.29 |
| 42.16 | 43.59 | 44.71 | 45.98 |
| 46.26 | 47.74 | 48.89 | 50.19 |

*CONSTANTE K:* 1384087.785
*DESIREZ-VOUS UTILISER LES DONNEES ORIGINALES?*
*OUI*
*FIN DE L'EXECUTION DE LA PROCEDURE D'ENTREE DU VECTEUR DE DONNEES.*

Après avoir entré les données, l'utilisateur sera appelé à choisir le travail qu'il désire faire exécuter. La liste des tâches effectuées par ANOMHI sera imprimée. Chacune est numérotée. Pour obtenir l'une d'entre elles, il suffit d'inscrire le numéro correspondant à la tâche désirée.

EXEMPLE 1.5: LISTE DES TRAVAUX DISPONIBLES

```
QUEL TRAVAIL DESIREZ-VOUS FAIRE EXECUTER?
1)  UNE ANALYSE GENERALE D'UN TABLEAU DE CONTINGENCE MULTIVARIE?
2)  UNE ANALYSE DE REGRESSION?
3)  L'EVALUATION DES FREQUENCES ESTIMEES PAR UN OU DES MODELES NON-SATURES?
4)  LA REPRODUCTION ET LA NORMALISATION DU (DES) TABLEAU(X) DE FREQUENCES
    EVALUE(S) PAR UN MODELE?
5)  LE CALCUL DES COEFFICIENTS DE DETERMINATION MULTIPLE?
6)  LE CALCUL DE(S) COEFFICIENT(S) DE DETERMINATION PARTIELLE?
7)  LE CALCUL DE(S) COEFFICIENTS(S) D'EFFICACITE DE LA REDUCTION DE
    CATEGORIES?
(INSCRIVEZ LE NUMERO DE LA QUESTION QUI CORRESPOND AU TRAVAIL DESIRE)
[]:
```

### B) Les sous-routines de calcul des coefficients d'association

Deux sous-routines sont disponibles selon les besoins: 1) le calcul des coefficients $\lambda$ d'un modèle d'analyse générale (L.A. Goodman, 1972a) et 2) le calcul des coefficients $\Lambda$ d'un modèle de régression (L.A. Goodman, 1972b). Dans chacun des deux cas, il est possible de traiter plusieurs échantillons. Il faudra alors s'assurer que les données pertinentes à chacun des échantillons soient inscrites en colonnes. Pour les fins de l'entrée des données, l'ensemble des échantillons est considéré comme la dernière variable dont chacune des catégories représente un échantillon (cf. tableau 25 où

les échantillons seront entrés comme la variable D). Une question spécifique sera demandée en regard du traitement de plusieurs échantillons (L.A. Goodman, 1973).

Le programme d'analyse générale permet de considérer qu'une variable et une seule possède des catégories équidistantes (L.A. Goodman, 1971). Si le cas se présente, cette variable sera identifiée, à la question pertinente, par le caractère qui la représente.

*EXEMPLE 2.1: QUESTIONS DU PROGRAMME D'ANALYSE GENERALE*

```
***ANALYSE GENERALE D'UN TABLEAU DE FREQUENCES***
TRAITEZ-VOUS UN SEUL ECHANTILLON?
OUI
LES CATEGORIES D'UNE VARIABLE SONT-ELLES EQUIDISTANTES?
OUI
QUELLE EST CETTE VARIABLE?
D
```

Le programme d'analyse de régression permet d'utiliser une ou plusieurs variables dépendantes. L'identification de ces variables se fait à l'aide du caractère qui les représentent. Si plusieurs variables doivent être identifiées, aucun espace entre les caractères ne doit apparaître.

*EXEMPLE 2.2: QUESTIONS DU PROGRAMME D'ANALYSE DE REGRESSION*

```
***ANALYSE DE REGRESSION***
QUELLE(S) EST (SONT) LA (LES) VARIABLE(S) DEPENDANTE(S)?
C
TRAITEZ-VOUS UN SEUL ECHANTILLON?
OUI
```

Les coefficients sont identifiés par les caractères qui représentent les variables et par des numéros qui représentent les catégories. De façon générale, seuls les coefficients nécessaires à l'identification complète des fréquences apparaissent. Dans le cas du traitement de plusieurs échantillons, les coefficients de différences entre tous les échantillons sont inclus pour faciliter l'analyse.

EXEMPLE 2.3: SORTIE DU SOUS PROGRAMME D'ANALYSE DE REGRESSION AVEC C COMME VARIABLE DEPENDANTE

| PARAMETRES RELATIONNELS | TAU | LAMBDA | LAMBDA STANDARDISE |
|---|---|---|---|
| C2 | 0.8017 | - 0.2211 | - 3.9875 |
| C3 | 0.7239 | - 0.3232 | - 6.0895 |
| A1C2 | 0.8512 | - 0.1612 | - 2.9065 |
| A1C3 | 0.8107 | - 0.2099 | - 3.9549 |
| B1C2 | 0.9234 | - 0.0796 | - 1.4386 |
| B1C3 | 0.8775 | - 0.1307 | - 2.4623 |
| D1C2 | 0.9560 | - 0.0450 | - 0.4162 |
| D1C3 | 0.9318 | - 0.0706 | - 0.6824 |
| D2C2 | 0.9148 | - 0.0891 | - 0.9108 |
| D2C3 | 0.9400 | - 0.0619 | - 0.6581 |
| D3C2 | 1.0578 | 0.0561 | 0.6215 |
| D3C3 | 1.0529 | 0.0516 | 0.5981 |
| A1B1C2 | 0.9332 | - 0.0692 | - 2.1078 |
| A1B1C3 | 0.8942 | - 0.1119 | - 2.1078 |
| A1D1C2 | 0.9587 | - 0.0422 | - 0.3900 |
| A1D1C3 | 0.9366 | - 0.0655 | - 0.6334 |
| A1D2C2 | 0.9158 | - 0.0603 | - 0.6416 |
| A1D2C3 | 0.9415 | 0.0552 | 0.6106 |
| A1D3C2 | 1.0587 | 0.0498 | 0.5778 |
| A1D3C3 | 1.0511 | - 0.0937 | - 0.8666 |
| B1D1C2 | 0.9105 | 0.0794 | 0.8116 |
| B1D2C2 | 1.0826 | 0.0463 | 0.4922 |
| B1D2C3 | 1.0474 | - 0.0018 | - 0.0197 |
| B1D3C2 | 0.9982 | 0.0126 | 0.1459 |
| B1D3C3 | 1.0127 | - 0.0927 | - 0.8576 |
| A1B1D1C2 | 0.9114 | - 0.0926 | - 0.8950 |
| A1B1D1C3 | 0.9115 | 0.0797 | 0.8145 |
| A1B1D2C2 | 1.0829 | 0.0488 | 0.4974 |
| A1B1D2C3 | 1.0479 | - 0.0021 | - 0.0235 |
| A1B1D3C2 | 0.9979 | 0.0120 | 0.1391 |
| A1B1D3C3 | 1.0121 | | |

C) *L'estimation des fréquences par un ou des modèles non saturés*

La sous-routine d'estimation des fréquences permet de traiter trois types d'hypothèses qui requièrent un traitement spécifique:
1) des hypothèses d'indépendance régionale (F. Béland, 1978),
2) des hypothèses de quasi-indépendance (L.A. Goodman, 1968) et
3) des hypothèses d'indépendance généralisée (Y.M.M. Bishop et al. 1975).

Le troisième type d'hypothèse sera probablement le plus utilisé. Il sera traité en premier lieu.

a) Hypothèse d'indépendance généralisée

En ce cas, l'utilisateur aura répondu négativement aux deux premières questions demandées par le sous-programme. La question suivante lui demandera de spécifier le critère d'ajustement itératif pour chacun des modèles qu'il entend estimé. Il est important que chacun des modèles soit identifié par son propre critère d'ajustement. Ce critère déterminera le degré de précision des estimés de fréquences sous un modèle particulier. Les valeurs .01 et .001 produisent des résultats satisfaisants. Au cours de l'exécution du programme, il sera notifié si un nombre trop grand d'itérations sont nécessaires pour produire le degré de précision désiré. L'utilisateur pourra alors décider d'interrompre ou de laisser poursuivre le processus itératif.

Les modèles s'inscrivent à l'aide des caractères qui représentent les variables. Les hypothèses dont il est question cherchent à réduire le nombre d'interactions nécessaires pour expliquer la variation du tableau de contingence représenté par le Khi deux d'indépendance ou le Khi deux d'équiprobabilité. Dans le cas d'un tableau à quatre variables identifiées chacune par les caractères *A*, *B*, *C*, et *D*, l'interaction conjointe des quatre variables, représentée par la conjonction des quatre variables *ABCD*, explique évidemment l'ensemble de la variance. Il s'agit donc d'associer les variables en entités satisfaisantes selon un modèle théorique pré-déterminé ou selon une volonté empirique de réduction pour étude ultérieure. Les assemblages de variables qui représentent des interactions ou des associations seront spécifiées par l'impression successive des caractères[1]. Ainsi une association entre les variables *A* et *C* sera écrite *AC*. Puisqu'il s'agira souvent de modèles multivariés où plusieurs types d'associations et d'interactions seront vérifiés par un même modèle, une virgule séparera chacune des associations et interactions pertinentes. Ainsi à l'association entre *A* et *C* peut s'ajouter une interaction entre *A*, *B* et *D*. Le modèle s'écrira: *AC,ABD*. L'indépendance entre chacune de ces variables s'écrit: *A,B,C,D*. L'hypothèse d'une association entre toutes les variables du tableau 25 deux à deux s'écrit: *AB,AC,AD,BC,BD,CD*. Ne pas inscrire de point à la fin des modèles. L'utilisateur ne devra pas laisser d'espace

---

1. Sauf l'hypothèse d'équiprobabilité qui est représentée par ○, soit le caractère obtenu par la touche supérieure de *O*.

vide entre aucun des caractères des modèles. Il devra inscrire un seul modèle par ligne. Une procédure de correction est prévue pour l'entrée de l'ensemble des commandes. Si l'une des fréquences a pour valeur estimée le nombre zéro à un moment quelconque du processus itératif d'estimation un message sera imprimé pour signaler la nécessité de corriger le nombre de degrés de liberté. Enfin les statistiques fournies sont: 1) les fréquences estimées; 2) le Khi deux usuel, 3) le Khi deux du rapport de vraisemblance et 4) les déviations des fréquences estimées des fréquences observées. Les statistiques 1 et 4 sont optionnelles.

*EXEMPLE 3.1:* *QUESTIONS ET SORTIES DU PROGRAMME D'EVALUATION DES FREQUENCES ESTIMEES*

```
***EVALUATION DES FREQUENCES ESTIMEES PAR UN OU DES MODELES NON-SATURES***
VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) D'INDEPENDANCE REGIONALE?
NON
VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) DE QUASI-INDEPENDANCE?
NON
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
□:
     .01   .01   .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
NON
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ
TRAITER:
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
A,B,C,D
AB,AC,AD,BC,BD,CD
AB,ABC
VERIFIEZ VOS ENTREES, AVEZ-VOUS COMMIS DES ERREURS?
OUI
RECOMMENCEZ ET CETTE FOIS FAITES ATTENTION
VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) DE QUASI-INDEPENDANCE?
NON
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
□:
      .01    .01    .01
```

```
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
OUI
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ
TRAITER:
A,B,C,D
AB,AC,AD,BC,BD,CD
ABD,ACD
VERIFIEZ VOS ENTREES. AVEZ-VOUS COMMIS DES ERREURS?
NON
 MODELE VERIFIE:  A,B,C,D
FREQUENCES ESTIMEES:
    7.28    7.83    7.92    8.23
    9.03    9.45    9.82   10.20
   10.32   10.80   11.21   11.66
   12.03   12.59   13.07   13.59
   14.90   15.60   16.20   16.84
   17.03   17.83   18.51   19.25
   21.20   22.20   23.04   23.96
   26.27   27.51   28.55   29.70
   30.02   31.44   32.63   33.94
   35.00   36.65   38.04   39.56
   43.47   45.41   47.14   49.02
   49.56   51.90   53.87   56.03
KHI DEUX USUEL:  42
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  48.02
DEVIATIONS DES ESTIMES:
   -8.28   -5.63   -4.92   -4.23
   -4.03   -3.45   -2.81   -2.20
   -1.32   -0.80   -0.21    0.34
    0.97   -4.59    1.93    2.41
    2.10    9.40    2.80    3.16
    3.97    4.17    4.49    4.75
    3.80    3.80    3.96    4.04
    2.73    2.49    2.45    2.30
    2.98    2.56    2.37    2.06
    2.00    2.35    0.96    0.44
   -2.37   -3.41   -4.14   -5.02
   -4.56   -5.90   -6.87   -8.03
 MODELE VERIFIE:  AB,AC,AD,BC,BD,CD
FREQUENCES ESTIMEES:
    3.07    3.22    3.91    4.34
    5.64    6.76    8.98    7.68
    7.82    8.66    9.53   10.40
   10.52   10.74   12.58   13.62
   17.46   20.39   20.34   21.78
   21.51   23.22   24.66   26.20
   23.68   23.15   26.59   28.04
   27.80   31.07   30.40   31.68
   34.00   35.14   36.59   37.86
   38.74   36.89   40.92   41.99
   41.11   44.77   42.28   42.89
   44.68   44.98   45.22   45.54
```

b) Hypothèse de quasi-indépendance

Il est possible que des cellules d'un tableau de contingence contiennent des valeurs égales à zéro; ou encore qu'une seule partie d'un tableau intéresse le chercheur. Il s'agit alors de traiter d'hypothèses de quasi-indépendance (L.A. Goodman, 1968). Une procédure d'identification des cellules vides est prévue si une ou des hypothèses de quasi-indépendance veulent être vérifiées. Un ou des vecteurs de 1 et de 0 appelés des delta-contrastes seront inscrits sous la question appropriée de la même manière et selon le même ordre que les fréquences du tableau original. La valeur 1 indique que la cellule reproduit la fréquence originale, la valeur 0 indique une cellule vide. Si plusieurs hypothèses de quasi-indépendance avec des structures de delta-contrastes différentes sont vérifiées, il suffit d'inscrire autant de vecteurs delta-contrastes qu'il y a d'hypothèses les uns à la suite des autres sans aucune identification. Si une seule structure est vérifiée pour plusieurs hypothèses, un seul vecteur peut être entré. Le reste des procédures suit le cas des hypothèses d'indépendance généralisée.

*EXEMPLE 3.2: ENTREE DU VECTEUR DES DELTA-CONTRASTES*

```
***EVALUATION DES FREQUENCES ESTIMEES PAR UN OU DES MODELES NON-SATURES***

VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) D'INDEPENDANCE REGIONALE?
NON
VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) DE QUASI-INDEPENDANCE?
OUI
ENTREZ LE(S) VECTEUR(S) DE DELTA-CONTRASTES:
□:
     1    1    1    1
```

```
[]:
    1   1   1   1
[]:
    1   1   0   0
[]:
    1   1   1   1
[]:
    1   1   1   1
[]:
    1   1   0   0
[]:
    1   1   1   1
[]:
    1   1   1   1
[]:
    1   1   0   0
[]:
    1   1   1   1
[]:
    1   1   1   1
[]:
    1   1   0   0
[]:
    10
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
[]:
    .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
OUI
INSTRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
A,B,C,D
VERIFIEZ VOS ENTREES.  AVEZ-VOUS COMMIS DES ERREURS?
NON
```

c) Hypothèses d'indépendance régionale

Les hypothèses d'indépendance régionale permettent de vérifier la légitimité du regroupement de catégories d'une ou de plusieurs variables (F. Béland, 1978). Essentiellement le modèle log-linéaire de l'hypothèse de réduction est construit à partir de deux types de paramètres: ceux maintenus du tableau original et ceux qui appartiennent au tableau réduit produit par le regroupement des ca-

tégories. En effet, la réduction des catégories des variables peut se faire sur des paramètres d'association spécifiques tels que définient par une hypothèse de regroupement quelconque. Conceptuellement il est nécessaire dans une première étape de spécifier les paramètres d'association à conserver du tableau original et dans une deuxième étape, le modèle log-linéaire du tableau réduit est identifié.

Pour obtenir la sous-routine d'estimation des fréquences sous une ou des hypothèses de regroupement de catégories, c'est-à-dire d'indépendance régionale, il suffit de répondre *OUI* à la question: *VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) D'INDEPENDANCE REGIONALE?* posée au début du programme d'estimation des fréquences sous des modèles log-linéaires. Dans un premier moment, il sera nécessaire de spécifier les paramètres retenus du tableau original sous la forme habituelle de modèle log-linéaire. Ces modèles sont dits d'indépendance généralisée pour employer l'expression de H.H. Ku et S. Kulback (1968). Le deuxième moment consiste à identifier les regroupements de catégories. Plusieurs hypothèses de regroupement peuvent être vérifiées, il suffit d'en indiquer le nombre à la question pertinente pour voir la sous-routine assurer un retour automatique à la procédure d'identification des regroupements. Les variables sont indiquées par la méthode habituelle employée par ANOMII: le caractère respectif qui les représente. Une question demandera d'indiquer les groupes de catégories amalgamées. Seules les catégories soumises à un regroupement doivent apparaître. Si plusieurs regroupements sont faits de différentes catégories d'une même variable, une virgule sépare les groupes. Ainsi pour une variable ayant

quatre catégories dont les classes 1 et 3 forment un premier groupe et les classes 2 et 4 en forment un second, ces deux groupes seront identifiés de la manière suivante après la commande *INSCRIVEZ LES GROUPES DE CATEGORIES*: 1 3,2 4. Un espace doit séparer chacune des catégories tandis qu'une virgule sépare les groupes de catégories.

L'entrée d'un zéro à la question: *QUELLE EST LA VARIABLE TRAITEE?* termine le procédé. Le troisième moment prévoit l'entrée de(s) modèle(s) log-linéaire(s) du tableau réduit. Chacune des entrées du modèle se fait selon la procédure habituelle. Une seule exception à cette règle: si une seule hypothèse d'indépendance généralisée est entrée, autant de modèles log-linéaires du tableau réduit qu'il en est désiré peuvent être vérifiés. Mais si plusieurs hypothèses d'indépendance généralisée sont possibles, une utilisation judicieuse du code *IDEM* qui répète le modèle qui le précède facilitera la procédure et économisera du temps d'ordinateur. (cf. l'exemple 3.3).

La sous-routine est souple en ce qu'elle permet de vérifier des modèles log-linéaires non testés lors d'une première utilisation, de même que d'utiliser de nouvelles hypothèses d'indépendance généralisée avec le même regroupement. Lors d'un changement d'hypothèse de regroupement, les hypothèses d'indépendance généralisée employées lors de la précédente utilisation de la sous-routine peuvent être conservées. Le temps d'ordinateur est ainsi économisé.

```
KHI DEUX USUEL:  8.45
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  8.92
DEVIATIONS DES ESTIMES:
   -2.07   -1.22   -0.91   -0.34
   -0.64   -0.76    0.02    0.32
    1.18    1.34    1.47    1.60
    2.48   -2.74    2.42    2.38
   -0.46    4.61   -1.34   -1.76
   -0.51   -1.22   -1.66   -2.20
    1.32    2.85    0.41   -0.04
    1.20   -1.07    0.60    0.32
   -1.00   -1.14   -1.59   -1.86
   -1.74    1.11   -1.92   -1.99
   -0.11   -2.77    0.72    1.11
    0.32    1.02    1.78    2.46

MODELE VERIFIE:  AB,ACD
FREQUENCES ESTIMES:
    3.63    2.59    4.66    5.18
    5.70    8.03    6.74    7.26
    7.77    8.29    8.81    7.26
   10.37    7.41   13.34   14.82
   16.30   22.97   19.26   20.74
   22.23   23.71   25.19   26.67
   25.90   26.74   27.58   28.41
   29.25   26.74   27.58   31.75
   32.59   33.42   34.26   35.10
   36.10   37.26   38.42   39.59
   40.75   41.92   43.08   44.25
   45.41   46.58   47.74   48.90
KHI DEUX USUEL:  7.57
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  8.33

DEVIATIONS DES ESTIMES:
   -2.64   -0.59   -1.66   -1.18
   -0.70   -2.03    0.26    0.74
    1.23    1.71    2.19    2.67
    2.63    0.59    1.66    1.18
    0.70    2.03   -0.26   -0.74
   -1.23   -1.71   -2.19   -2.67
   -0.90   -0.74   -0.58   -0.41
   -0.25   -0.08    0.08    0.25
    0.41    0.58    0.74    0.90
    0.90    0.74    0.58    0.41
    0.25    0.08   -0.08   -0.25
   -0.41   -0.58   -0.74   -0.90
DESIREZ-VOUS ENTRER D'AUTRES DONNEES?
NON
DESIREZ-VOUS FAIRE EXECUTER UN AUTRE TRAVAIL?
OUI
(INSCRIVEZ LE NUMERO DE LA QUESTION QUI CORRESPOND AU TRAVAIL DESIRE)
[]:
     3
```

EXEMPLE 3.3: EMPLOI DE LA SOUS-ROUTINE D'ESTIMATION DES HYPOTHESES D'INDEPENDANCE REGIONALE

```
***EVALUATION DES FREQUENCES ESTIMEES PAR UN OU DES MODELES NON SATURES***

VOULEZ-VOUS VERIFIER UNE (DES) HYPOTHESE(S) D'INDEPENDANCE REGIONALE?
OUI
1)  HYPOTHESE(S) D'INDEPENDANCE GENERALISEE
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
[]:
     .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
NON
INSCRIVEZ LA (LES) MARGINALE (S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
A,B,C,D
VERIFIEZ VOS ENTREES. AVEZ-VOUS COMMIS DES ERREURS?
NON
MODELE VERIFIE:  A,B,C,D
KHI DEUX USUEL:  42
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  48.02

2)  HYPOTHESE(S) D'INDEPENDANCE REGIONALE
COMBIEN D'HYPOTHESES D'INDEPENDANCE REGIONALE VOULEZ-VOUS VERIFIER?
[]:
     2
HYPOTHESE REGIONALE NO. 1
QUELLE EST LA VARIABLE TRAITEE?
C
INSCRIVEZ LES GROUPES DE CATEGORIES:
1 2
QUELLE EST LA VARIABLE TRAITEE?
D
INSCRIVEZ LES GROUPES DE CATEGORIES:
1 2,3 4
QUELLE EST LA VARIABLE TRAITEE?
0
VERIFIEZ VOS ENTREES.  AVEZ-VOUS COMMIS DES ERREURS?
NON

STRUCTURE(S) D'HYPOTHESE DU TABLEAU REGIONALISE
SPECIFIEZ LE(S) CRITERES D'AJUSTEMENT ITERATIF:
[]:
      .01   .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
NON
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
```

```
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
ABC,ABD,ACD,BCD
ABCD
VERIFIEZ VOS ENTREES. AVEZ-VOUS COMMIS DES ERREURS?
NON
STRUCTURE D'HYPOTHESE DU TABLEAU REGIONALISE:  ABC,ABD,ACD,BCD
KHI DEUX USUEL:  12.22
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  12.93

STRUCTURE D'HYPOTHESE DU TABLEAU REGIONALISE:  ABCD
KHI DEUX USUEL:  12.1
KHI DEUX DU RAPPORT DE VRAISEMBLANCE: 12.78
VOULEZ-VOUS VERIFIER D'AUTRES MODELES AVEC LA MEME REGIONALISATION?
OUI
VOULEZ-VOUS VERIFIER LES MEMES MODELES D'INDEPENDANCE GENERALISEE?
NON
1)  HYPOTHESE(S) D'INDEPENDANCE GENERALISEE
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
[]:
     .01    .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
NON
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
ABC,ABD
AB,AC,AD,BC,BD
VERIFIEZ VOS ENTREES.  AVEZ-VOUS COMMIS DES ERREURS?
NON
MODELE VERIFIE:  ABC,ABD
KHI DEUX USUEL:  5.11
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  5.35

MODELE VERIFIE:  AB,AC,AD,BC,BD
KHI DEUX USUEL:  9.31
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  9.81

2)   HYPOTHESE(S) D'INDEPENDANCE REGIONALE

STRUCTURE(S) D'HYPOTHESE DU TABLEAU REGIONALISE
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
[]:
     .01    .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
NON
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
ABC,ABD,ACD,BCD
IDEM
VERIFIEZ VOS ENTREES.  AVEZ-VOUS COMMIS DES ERREURS?
NON
```

```
STRUCTURE D'HYPOTHESE DU TABLEAU REGIONALISE: ABC,ABD,ACD,BCD
KHI DEUX USUEL:  4.98
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  5.18

STRUCTURE D'HYPOTHESE DU TABLEAU REGIONALISE:  IDEM
KHI DEUX USUEL:  6.07
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  6.53

VOULEZ-VOUS VERIFIER D'AUTRES MODELES AVEC LA MEME REGIONALISATION?
NON
VOULEZ-VOUS VERIFIER LES MEMES MODELES D'INDEPENDANCE GENERALISEE?
OUI

HYPOTHESE REGIONALE NO. 2
QUELLE EST LA VARIABLE TRAITEE?
C
INSCRIVEZ LES GROUPES DE CATEGORIES:
1 2
QUELLE EST LA VARIABLE TRAITEE?
D
INSCRIVEZ LES GROUPES DE CATEGORIES:
1 3,2 4
QUELLE EST LA VARIABLE TRAITEE?
0
VERIFIEZ VOS ENTREES.  AVEZ-VOUS COMMIS DES ERREURS?
NON

STRUCTURE(S) D'HYPOTHESE DU TABLEAU REGIONALISE
SPECIFIEZ LE(S) CRITERE(S) D'AJUSTEMENT ITERATIF:
0:
     .01    .01
VOULEZ-VOUS L'IMPRESSION DES ESTIMES DES FREQUENCES?
NON
INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE VOUS VOULEZ TRAITER:
(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)
ABC,ABD,ACD,BCD
IDEM
VERIFIEZ VOS ENTREES.  AVEZ-VOUS COMMIS DES ERREURS?
NON
STRUCTURE D'HYPOTHESE DU TABLEAU REGIONALISE:  ABC,ABD,ACD,BCD
KHI DEUX USUEL:  5.02
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  5.29

STRUCTURE D'HYPOTHESE DU TABLEAU REGIONALISE:  IDEM
KHI DEUX USUEL:  6.45
KHI DEUX DU RAPPORT DE VRAISEMBLANCE:  6.98
VOULEZ-VOUS VERIFIER D'AUTRES MODELES AVEC LA MEME REGIONALISATION?
NON
```

D) La normalisation des fréquences

Les fréquences d'un échantillon sont normalisées soit en rapport avec les valeurs des marges ou des interactions d'un autre échantillon où les mêmes variables sont observées, soit pour isoler l'effet d'interaction d'ordre supérieur des effets d'ordre inférieur ou des effets des marges (F. Mosteller, 1968).

ANOMMI permet d'imposer à un échantillon des valeurs de marges ou d'interaction d'un autre échantillon dans le cas d'un tableau complet seulement (Y.M.M. Bishop et al, 1975:101). Seulement si le tableau de contingence ne contient pas de zéros structuraux, il sera demandé si cette procédure est désirée. Si oui, ANOMMI considérera que chacun des effets ou des marges à ajuster forme un ensemble. Le nombre d'ensemble sera demandé. Si quatre variables nommées $A$, $B$, $C$ et $D$ d'un premier échantillon sont ajustées aux marges $C$ et $D$ et à l'interaction $AB$ d'un autre échantillon, trois ensembles seraient donc construits. Une variable ne peut appartenir qu'à un seul ensemble et chaque ensemble doit s'additionner à la même somme. Une question spécifique permet d'identifier les ensembles, une autre d'entrer les valeurs correspondantes. Ces valeurs doivent être considérées comme des tableaux ordinaires et inscrites en conformité avec les exigences d'ANOMMI à cet égard.

Les fréquences normalisées sont alors fournies par le programme et la réponse à une question permet à l'utilisateur de poursuivre ses travaux soit avec les fréquences originales, soit avec les fréquences normalisées.

La seconde phase d'utilisation de cette sous-routine permet de normaliser les fréquences d'un tableau de contingence pour isoler les interactions des effets secondaires. Si un modèle de régression est traité, la ou les variables dépendantes sont identifiées, puis la normalisation pourra s'appliquer aux interactions, alors sans intérêt, des variables indépendantes entre elles. La normalisation s'applique à la ou aux variables dépendantes et à l'ensemble constitué par les variables indépendantes. L'interaction seule entre les premières et les dernières est donc identifiée.

La normalisation se fait à la fois sur les fréquences et sur les proportions. Dans le cas d'un modèle général, la somme du tableau des proportions normalisées sera de un. Dans le cas d'un modèle de régression, la somme sur les rangées du tableau sera égale à l'unité pour permettre une comparaison des effets des variables indépendantes sur chacune des valeurs de la ou des variables dépendantes. Si le modèle de régression comprote plusieurs variables dépendantes, celles-ci paraîtront en colonne sous la forme d'une seule variable, la première étant fixée, la seconde variera et ainsi de suite (tableau 26). S'il n'y a qu'une seule variable dépendante placée en colonne dans le tableau original, il n'y aura pas de remaniement du tableau. Un seul sera alors présenté avec la somme des proportions sur la variable dépendante égale à l'unité.

Des interactions spécifiques peuvent être isolées. Une question permet d'identifier ces interactions. La forme d'entrée de ces interactions est la même que celle de l'entrée des modèles dans le pro-

gramme d'estimation des fréquences. Autant de paramètres que désirés peuvent être traités. Les résultats de la normalisation seront donnés pour chacun d'entre eux.

Tableau 26

Tableau 25 remanié lorsque *A* et *B* sont des variables dépendantes

| | | *A* | 1 | | 2 | |
|---|---|---|---|---|---|---|
| | | *B* | 1 | 2 | 1 | 2 |
| *C* | *D* | | | | | |
| 1 | 1 | | 1111 | 1211 | 2111 | 2211 |
| 1 | 2 | | 1112 | 1212 | 2112 | 2212 |
| 1 | 3 | | 1113 | 1213 | 2113 | 2213 |
| 1 | 4 | | 1114 | 1214 | 2114 | 2214 |
| 2 | 1 | | 1121 | 1221 | 2121 | 2221 |
| 2 | 2 | | 1122 | 1222 | 2122 | 2222 |
| 2 | 3 | | 1123 | 1223 | 2123 | 2223 |
| 2 | 4 | | 1124 | 1224 | 2124 | 2224 |
| 3 | 1 | | 1131 | 1231 | 2131 | 2231 |
| 3 | 2 | | 1132 | 1232 | 2132 | 2232 |
| 3 | 3 | | 1133 | 1233 | 2133 | 2233 |
| 3 | 4 | | 1134 | 1234 | 2134 | 2234 |

*EXEMPLE 4. UTILISATION DU SOUS-PROGRAMME DE NORMALISATION DES FREQUENCES LORSQUE LE TABLEAU EST COMPLET*

```
***PROCEDURE ITERATIVE DE NORMALISATION DES FREQUENCES***
VOTRE TABLEAU COMPTE-T-IL DES ZEROS STRUCTURAUX?
NON
SPECIFIEZ LE CRITERE D'AJUSTEMENT ITERATIF:
□:
      .01
```

```
VOULEZ-VOUS AJUSTER VOS FREQUENCES A CELLES D'UN AUTRE ECHANTILLON?
OUI

COMBIEN D'ENSEMBLES DE MARGINALES DESIREZ-VOUS ENTRER?
[]:
    3
INSCRIVEZ UN ENSEMBLE DE MARGINALE(S):
AB
INSCRIVEZ LES VALEURS DES FREQUENCES DES MARGINALES:
[]:
    25 30 35 40
[]:
    10
INSCRIVEZ UN ENSEMBLE DE MARGINALE(S):
C
INSCRIVEZ LES VALEURS DES FREQUENCES DES MARGINALES:
[]:
    40 50 40
[]:
    10
INSCRIVEZ UN ENSEMBLE DE MARGINALE(S):
D
INSCRIVEZ LES VALEURS DES FREQUENCES DES MARGINALES:
[]:
    30 20 50 30
[]:
    10
AVEZ-VOUS COMMIS DES ERREURS?
NON

LES FREQUENCES NORMALISEES:
    0.43    0.53    1.87    1.41
    1.97    1.47    4.04    2.81
    2.32    1.61    4.16    2.57
    2.19    0.84    3.71    2.24
    2.65    2.43    4.34    2.59
    2.15    1.40    3.45    2.04
    2.95    1.91    4.67    2.74
    3.16    2.04    4.96    2.90
    2.36    1.51    2.67    2.14
    3.56    2.27    5.50    3.19
    3.65    2.33    5.61    3.25
    2.62    1.67    4.02    2.32
VOULEZ-VOUS CONTINUER VOS TRAVAUX AVEC LES FREQUENCES NORMALISEES?
OUI

***VOTRE VECTEUR DE DONNEES EST MAINTENANT CONSTITUE DES FREQUENCES
   NORMALISEES***
VOULEZ-VOUS NORMALISER LES TABLEAUX DES FREQUENCES EN INTERACTION?
OUI
?TRAITEZ-VOUS UN MODELE DE REGRESSION?
OUI
```

QUELLE(S) EST (SONT) LA (LES) VARIABLE(S) DEPENDANTE(S)?
AB
VOULEZ-VOUS NORMALISER LES INTERACTIONS DES VARIABLES INDEPENDANTES?
OUI

ATTENTION, DEUX TABLEAUX SERONT PRESENTES POUR CHAQUE SORTIE.
DANS LE PREMIER, L'ORDRE DES VARIABLES SERA CDAB PLUTOT QUE ABCD.
LA (LES) VARIABLE(S) DEPENDANTE(S) AB APPARAITRA(ONT) EN COLONNES.
LE SECOND TABLEAU RESPECTERA LA FACON ORIGINALE D'ENTREE DES DONNEES.

LES FREQUENCES NORMALISEES:
PREMIER TABLEAU:

| | | | |
|---|---|---|---|
| 0.73 | 3.05 | 3.44 | 3.62 |
| 1.49 | 1.91 | 3.64 | 3.79 |
| 1.79 | 2.88 | 3.04 | 3.13 |
| 2.19 | 2.82 | 2.89 | 2.94 |
| 2.52 | 2.77 | 2.65 | 2.78 |
| 2.56 | 3.44 | 2.42 | 2.41 |
| 3.06 | 2.68 | 2.56 | 2.53 |
| 3.28 | 2.64 | 2.48 | 2.43 |
| 3.48 | 2.61 | 2.41 | 2.33 |
| 3.65 | 2.59 | 2.34 | 2.25 |
| 3.80 | 2.56 | 2.28 | 2.18 |
| 3.94 | 2.54 | 2.23 | 2.12 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 0.73 | 1.49 | 1.79 | 2.19 |
| 2.52 | 2.56 | 3.06 | 3.28 |
| 3.48 | 3.65 | 3.80 | 3.94 |
| 3.05 | 1.91 | 2.88 | 2.82 |
| 2.77 | 3.44 | 2.68 | 2.64 |
| 2.61 | 2.59 | 2.56 | 2.54 |
| 3.44 | 3.64 | 3.04 | 2.89 |
| 2.76 | 2.42 | 2.56 | 2.48 |
| 2.41 | 2.34 | 2.28 | 2.23 |
| 3.62 | 3.79 | 3.12 | 2.94 |
| 2.78 | 2.41 | 2.53 | 2.43 |
| 2.33 | 2.25 | 2.18 | 2.12 |

LES PROPORTIONS NORMALISEES:
PREMIER TABLEAU:

| | | | |
|---|---|---|---|
| 0.067 | 0.202 | 0.317 | 0.334 |
| 0.137 | 0.177 | 0.336 | 0.350 |
| 0.165 | 0.266 | 0.281 | 0.288 |
| 0.202 | 0.260 | 0.167 | 0.271 |
| 0.233 | 0.255 | 0.255 | 0.257 |
| 0.237 | 0.319 | 0.223 | 0.222 |
| 0.283 | 0.247 | 0.236 | 0.211 |
| 0.303 | 0.244 | 0.229 | 0.224 |
| 0.321 | 0.241 | 0.222 | 0.216 |
| 0.337 | 0.239 | 0.216 | 0.208 |
| 0.351 | 0.237 | 0.211 | 0.201 |
| 0.364 | 0.235 | 0.206 | 0.195 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 0.006 | 0.011 | 0.014 | 0.017 |
| 0.019 | 0.020 | 0.024 | 0.025 |
| 0.027 | 0.028 | 0.029 | 0.030 |
| 0.023 | 0.015 | 0.022 | 0.022 |
| 0.021 | 0.026 | 0.021 | 0.020 |
| 0.020 | 0.020 | 0.020 | 0.020 |
| 0.026 | 0.028 | 0.023 | 0.022 |
| 0.021 | 0.019 | 0.020 | 0.019 |
| 0.019 | 0.018 | 0.018 | 0.017 |
| 0.028 | 0.029 | 0.024 | 0.023 |
| 0.021 | 0.019 | 0.019 | 0.019 |
| 0.018 | 0.017 | 0.017 | 0.016 |

VOULEZ-VOUS NORMALISER LES TABLEAUX DES PARAMETRES D'UN MODELE?
OUI
INSCRIVEZ LE(S) PARAMETRE(S) D'INTERACTION DU MODELE:
AC,BC,AD,BD,ABD

PARAMETRE CA

ATTENTION, DEUX TABLEAUX SEPONT PRESENTES POUR CHAQUE SROTIE.
DANS LE PREMIER, L'ORDRE DES VARIABLES SERA CA PLUTOT QUE AC.
LA (LES) VARIABLE(S) DEPENDANTE(S) A APPARAITRA(ONT) EN COLONNES.
LE SECOND TABLEAU RESPECTERA LA FACON ORIGINALE D'ENTREE DES DONNEES.

LES FREQUENCES NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 19.55 | 23.79 |
| 22.75 | 20.59 |
| 22.71 | 20.63 |

DEUXIEME TABLEAU:

| | | |
|---|---|---|
| 19.55 | 22.75 | 22.71 |
| 23.79 | 20.59 | 20.63 |

LES PROPORTIONS NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 0.451 | 0.549 |
| 0.525 | 0.475 |
| 0.524 | 0.476 |

DEUXIEME TABLEAU:

| | | |
|---|---|---|
| 0.150 | 0.175 | 0.175 |
| 0.183 | 0.158 | 0.159 |

LES FREQUENCES OBSERVEES:
PREMIER TABLEAU:

| | |
|---|---|
| 11.35 | 22.62 |
| 15.68 | 27.68 |
| 19.97 | 29.70 |

DEUXIEME TABLEAU:

| | | |
|---|---|---|
| 11.35 | 18.68 | 19.97 |
| 22.62 | 27.88 | 29.70 |

PARAMETRE CB

ATTENTION, DEUX TABLEAUX SERONT PRESENTES POUR CHAQUE SORTIE.
DANS LE PREMIER, L'ORDRE DES VARIABLES SERA CB PLUTOT QUE BC.
LA (LES) VARIABLE(S) DEPENDANTE(S) B APPARAITRE(ONT) EN COLONNES.
LE SECOND TABLEAU RESPECTERA LE PACON ORIGINALE D'ENTREE DES DONNEES.

LES FREQUENCES NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 21.41 | 21.92 |
| 21.50 | 21.83 |
| 22.09 | 21.24 |

DEUXIEME TABLEAU:

| | | |
|---|---|---|
| 21.31 | 21.50 | 22.09 |
| 21.92 | 21.83 | 21.24 |

LES PROPORTIONS NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 0.494 | 0.506 |
| 0.496 | 0.504 |
| 0.510 | 0.490 |

DEUXIEME TABLEAU:

| | | |
|---|---|---|
| 0.165 | 0.165 | 0.170 |
| 0.169 | 0.168 | 0.163 |

LES FREQUENCES OBSERVEES:
PREMIER TABLEAU:

| | |
|---|---|
| 20.69 | 13.28 |
| 28.32 | 19.03 |
| 30.99 | 18.69 |

DEUXIEME TABLEAU:

| | | |
|---|---|---|
| 20.69 | 28.32 | 30.99 |
| 13.28 | 19.03 | 18.69 |

PARAMETRE DA

ATTENTION, DEUX TABLEAUX SERONT PRESENTES POUR CHAQUE SORTIE.
DANS LE PREMIER, L'ORDRE DES VARIABLES SERA DA PLUTOT QUE AD.
LA (LES) VARIABLE(S) DEPENDANTE(S) A APPARAITRA(ONT) EN COLONNES.
LE SECOND TABLEAU RESPECTERA LA PACON ORIGINALE D'ENTREE DES DONNEES.

LES FREQUENCES NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 16.17 | 16.33 |
| 16.13 | 16.37 |
| 16.47 | 16.03 |
| 16.22 | 16.28 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 16.17 | 16.13 | 16.47 | 16.22 |
| 16.33 | 16.37 | 16.03 | 16.28 |

LES PROPORTIONS NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 0.498 | 0.502 |
| 0.496 | 0.504 |
| 0.507 | 0.493 |
| 0.499 | 0.501 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 0.124 | 0.124 | 0.127 | 0.125 |
| 0.126 | 0.126 | 0.123 | 0.125 |

LES FREQUENCES OBSERVEES:
PREMIER TABLEAU:

| | |
|---|---|
| 9.90 | 16.02 |
| 12.02 | 20.83 |
| 14.76 | 23.00 |
| 12.52 | 20.14 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 9.90 | 12.02 | 14.76 | 12.52 |
| 16.02 | 20.83 | 23.00 | 20.14 |

PARAMETRE DB

ATTENTION, DEUX TABLEAUX SERONT PRESENTES POUR CHAQUE SORTIE.
DANS LE PREMIER, L'ORDRE DES VARIABLES SERA DB PLUTOT QUE BD.
LA (LES) VARIABLE(S) DEPENDANTE(S) B APPARAITRA(ONT) EN COLONNES.
LE SECOND TABLEAU RESPECTERA LA FACON ORIGINALE D'ENTREE DES DONNEES.

LES FREQUENCES NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 15.46 | 17.04 |
| 16.25 | 16.25 |
| 16.40 | 16.10 |
| 16.89 | 15.61 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 15.46 | 16.25 | 16.40 | 16.89 |
| 17.04 | 16.25 | 16.10 | 15.61 |

LES PROPORTIONS NORMALISEES:
PREMIER TABLEAU:

| | |
|---|---|
| 0.476 | 0.524 |
| 0.500 | 0.500 |
| 0.505 | 0.495 |
| 0.520 | 0.480 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 0.119 | 0.125 | 0.126 | 0.130 |
| 0.131 | 0.125 | 0.124 | 0.120 |

LES FREQUENCES OBSERVEES:
PREMIER TABLEAU:

| | |
|---|---|
| 15.32 | 10.61 |
| 20.66 | 12.99 |
| 23.36 | 14.40 |
| 20.66 | 12.01 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 15.32 | 20.66 | 23.36 | 20.66 |
| 10.61 | 12.99 | 14.40 | 12.01 |

PARAMETRE DAB

ATTENTION, DEUX TABLEAUX SERONT PRESENTES POUR CHAQUE SORTIE.
DANS LE PREMIER, L'ORDRE DES VARIABLES SERA DAB PLUTOT QUE ABD.
LA (LES) VARIABLE(S) DEPENDANTE(S) AB APPARAITRA(ONT) EN COLONNES.
LE SECOND TABLEAU RESPECTERA LA FACON ORIGINALE D'ENTREE DES DONNEES.

LES FREQUENCES NORMALISEES:
PREMIER TABLEAU:

| | | | |
|---|---|---|---|
| 7.79 | 8.46 | 8.46 | 7.79 |
| 8.14 | 8.11 | 8.11 | 8.14 |
| 8.15 | 8.10 | 8.10 | 8.15 |
| 8.42 | 7.83 | 7.83 | 8.42 |

DEUXIEME TABLEAU:

| | | | |
|---|---|---|---|
| 7.79 | 8.46 | 8.14 | 8.11 |
| 8.15 | 8.10 | 8.42 | 7.83 |
| 8.46 | 7.79 | 8.11 | 8.14 |
| 8.10 | 8.15 | 7.83 | 8.42 |

LES PROPORTIONS NORMALISEES:
PREMIER TABLEAU:

| | | | |
|---|---|---|---|
| 0.240 | 0.260 | 0.260 | 0.240 |
| 0.251 | 0.249 | 0.249 | 0.251 |
| 0.251 | 0.249 | 0.249 | 0.251 |
| 0.259 | 0.241 | 0.241 | 0.259 |

*DEUXIEME TABLEAU:*

| | | | |
|---|---|---|---|
| 0.060 | 0.065 | 0.063 | 0.062 |
| 0.063 | 0.062 | 0.065 | 0.060 |
| 0.065 | 0.060 | 0.062 | 0.063 |
| 0.062 | 0.063 | 0.060 | 0.065 |

*LES FREQUENCES OBSERVEES:*
*PREMIER TABLEAU:*

| | | | |
|---|---|---|---|
| 5.43 | 4.47 | 9.89 | 6.14 |
| 7.67 | 5.14 | 12.99 | 7.84 |
| 8.92 | 5.84 | 12.44 | 8.56 |
| 7.98 | 4.55 | 12.69 | 7.46 |

*DEUXIEME TABLEAU:*

| | | | |
|---|---|---|---|
| 5.43 | 9.89 | 7.67 | 12.99 |
| 8.92 | 14.44 | 7.98 | 12.69 |
| 4.47 | 6.14 | 5.14 | 7.84 |
| 5.84 | 8.56 | 4.55 | 7.46 |

*E) Les coefficients de détermination et d'efficacité*

Les coefficients de détermination multiple, multi-partielle et partielle permettent de mesurer la variation expliquée par un modèle par rapport 1o) à l'hypothèse d'équiprobabilité, 2o) d'indépendance totale et 3o) à un autre modèle qui inclut ou exclut un paramètre d'association (L.A. Goodman, 1972a). Les coefficients d'efficacité permettent de mesurer les effets de réduction de l'information causés par l'amalgamation des catégories selon une hypothèse d'indépendance régionale (F. Béland, 1978).

Ces coefficients sont obtenus par l'utilisation des programmes 5, 6 et 7. Les procédures sont simples; les valeurs des statistiques $X^2$ sont entrées avec leurs degrés de liberté à leur suite dans l'ordre imposé par les questions de chacun des programmes. Le programme de calcul des coefficients multiples et multi-partiels exigent que les

statistiques $X^2$ et les degrés de liberté des hypothèses d'équiprobabilité, d'indépendance totale et d'indépendance généralisée (modèle log-linéaire théorique ou empirique retenu) soient spécifiés.

*EXEMPLE 5.1.* *LES COEFFICIENTS DE DETERMINATION MULTIPLE ET MULTI-PARTIELLE*

```
***COEFFICIENTS DE DETERMINATION MULTIPLE***

ENTREZ VOS DONNEES:
A) HYPOTHESE D'EQUIPROBABILITE:
KHI DEUX: DEGRES DE LIBERTE:
?:
    438.02        47
B) HYPOTHESE D'INDEPENDANCE:
KHI DEUX: DEGRES DE LIBERTE:
?:
     48.02        40
C) HYPOTHESE A VERIFIER:
KHI DEUX:  DEGRES DE LIBERTE:
?:     8.92        23
       8.92        23
COEFFICIENT DE DETERMINATION MULTIPLE:  0.9585
COEFFICIENT DE DETERMINATION MULTI-PARTIELLE:  0.0769
```

Le programme de calcul des coefficients de détermination partielle exige qu'en premier lieu soit spécifié si le ou les paramètres d'association dont les contributions sont mesurées sont exclus ou inclus du modèle qui sert de contrôle. En effet, les statistiques de deux types d'hypothèses doivent être utilisées: 1) celles du modèle log-linéaire retenues comme satisfaisant d'un point de vue théorique ou empirique (hypothèse contrôle) et 2) celles des modèles qui excluent chacun un paramètre ou les modèles qui incluent chacun un paramètre. La fin des entrées du programme est signalée par deux zéros de la façon illustrée par l'exemple.

EXEMPLE 5.2. LES COEFFICIENTS DE DETERMINATION PARTIELLE DES PARAMETRES INCLUS DU MODELE AB,AC,AD,BC,BD,CD

```
***COEFFICIENT(S) DE DETERMINATION PARTIELLE***

VOULEZ-VOUS VERIFIER LA CONTRIBUTION D'UN (DE) PARAMETRE(S) INCLUS?
OUI
HYPOTHESE CONTROLE:
KHI DEUX:     DEGRES DE LIBERTE:
[]:
     8.92     23
HYPOTHESE(S) A VERIFIER:
KHI DEUX:     DEGRES DE LIBERTE:
[]:
     9.21     17
[]:
     9.22     20
[]:
    11.03     21
[]:
     9.85     20
[]:
    20.19     21
[]:
    35.02     22
[]:
     0         0
NO. DU PROBLEME:     COEFFICIENT(S) DE
                     DETERMINATION PARTIELLE:
      1.                  0.3279
      2.                  0.1507
      3.                  0.2616
      4.                  0.2125
      5.                  0.5966
      6.                  0.7584
```

Les coefficients d'efficacité sont obtenus par le programme no. 7. Les statistiques Khi deux et les degrés de liberté doivent être de la même façon que pour les deux programmes précédents. L'entrée de deux zéros signifie ici aussi la fin des inscriptions.

EXEMPLE 5.3. LES COEFFICIENTS D'EFFICACITE DE LA REDUCTION DE L'HYPOTHESE DE REGROUPEMENTS DES CATEGORIES 2 ET 3 DE LA VARIABLE C ET 1 2 PUIS 3 4 DE LA VARIABLE D SELON LES STRUCTURES D'HYPOTHESE DU TABLEAU REDUIT.

```
***COEFFICIENT(S) D'EFFICACITE D'UNE REDUCTION DE CATEGORIES***

HYPOTHESE D'INDEPENDANCE GENERALISEE:
KHI DEUX:     DEGRES DE LIBERTE:
[]:
    46.02            40
HYPOTHESE(S) D'INDEPENDANCE REGIONALE:
KHI DEUX:     DEGRES DE LIBERTE:
[]:
    12.93            29
[]:
    12.78            28
[]:
     0
NO. DU PROBLEME:      COEFFICIENT(S) D'EFFICACITE:
     1.                          0.7307
     2.                          0.7086
```

# BIBLIOGRAPHIE

BARTLETT, M.S. (1935). "Contingency Tables Interactions". Journal of the Royal Statistical Society Supplement, 2, pp. 248-252.

BELAND, F. (1978). L'indépendance régionale: la réduction des catégories dans les tableaux de contingence, Thèse de doctorat, Québec, Département de Sociologie, Université Laval.

BELAND, F., FORTIER, J.J. (1979). "Collapsible Contingency Tables", (à paraître, 1979).

BIRCH, M.W. (1963). "Maximum Likelihood in three-way Contingency Tables". Journal of the Royal Statistical Society, Series B, 25, pp. 220-233.

BISHOP, Y.M.M., FIENBERG, S.E. (1969). "Incomplete two-dimensional Contingency Tables". Biometrics, 24, pp. 119-128.

BISHOP, Y.M.M., FIENBERG, S., HOLLAND, P. (1974). Discrete Multivariate Analysis: Theory and Practice, Boston, M.I.T.

DAS, T. (1945). The Purums: An Old Kukitribe of Manipur, Calcutta, Université de Calcutta.

DAVIS, J.A. (1974). "Hierarchical Models for Significance Test in Multivariate Contingency Tables: An Exegesis of Goodman's Recent Papers", in, H.L. Costner, Sociological Methodology, 1973-74, San Francisco, Jossey-Bass, pp. 189-231.

FIENBERG, S.F., HOLLAND, P.W. (1973). "Simultaneous Estimation of Multinomial Cell Probabilities", J.A.S.A., 68, pp. 683-691.

GILLMAN, L., ROSE, A.J. (1974). APL, An Interactive Approach, New York, John Wiley.

GLASS, D.V. (1954). Social Mobility in Britain, Glencoe, Free Press.

GOODMAN, L.A. (1961). "Statistical Methods for the Movee-stayer Model", J.A.S.A., 56, pp. 841-868.

GOODMAN, L.A. (1963). "On Methods for Comparing Contingency Tables", J.R.S.S., A126, pp. 91-108.

GOODMAN, L.A. (1965). "On the Statistical Analysis of Mobility Tables", 70, pp. 564-585.

GOODMAN, L.A. (1968). "The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries", Journal of the American Statistical Association, 63, pp. 1071-1131.

GOODMAN, L.A. (1970). "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications". Journal of the American Statistical Association, 65, (No. 329), pp. 226-256.

GOODMAN, L.A. (1971). "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Buildings Models for Multiple Classifications". Technometrics, 13, pp. 33-61.

GOODMAN, L.A. (1972a). "A General Model for the Analysis of Surveys". American Journal of Sociology, 72, pp. 1035-1086.

GOODMAN, L.A. (1972b). "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables". A.S.R., 37, pp. 28-46.

GOODMAN, L.A. (1973). "Guided and Unguided Methods for Selecting Models for a Set of T Multidimensional Contingency Tables". J.A.S.A., 68, (No. 341), pp. 165-175.

GOODMAN, L.A., KRUSKALL, W.H. (1954). "Measures of Associations for Cross Classifications". J.A.S.A., 49, pp. 723-764.

GOODMAN, L.A., KRUSKALL, W.H. (1959). "Measures of Associations for Cross Classifications". J.A.S.A., 54, pp. 132-163.

GOODMAN, L.A., KRUSKALL, W.H. (1963). "Measures of Associations for Cross Classifications III. Approximate Sampling Theory". J.A.S.A., 58, pp. 310-364.

KIMBALL, D.W. (1954). "Short-cut Formulas for the Exact Partition of $\chi^2$ in Contingency Tables". Biometrics, 10, pp. 452-458.

KU, H.H., KULLBACK, S. (1968). "Interactions in Multidimentional Contingency Tables: An Information Theoretic Approach". Journal of Research of the National Bureau of Standards, 72b, pp. 154-199.

LAZARFELD, P. (1966). "L'algèbre des systèmes dichotomiques", in, R. Boudon et P. Lazarfeld, L'analyse empirique de la causalité, Paris, Mouton, pp. 255-275.

MOSTELLER, F. (1968). "Association and Estimation in Contingency Tables". Journal of the American Statistical Association, 63, pp. 1-28.

ROY, S.N., KASTENBAUM, M.A. (1956). "On the Hypothesis of no "interactions" in a multi-way contingency table". Annals of Mathematical Statistics, 27, pp. 749-757.

SMITH, K.W. (1976). "Marginal Standardization and Table Shrinking Aids in the Traditional Analysis of Contingency Tables". Social Forces, 54, pp. 669-693.

WHITE, H.C. (1963). "Cause and Effect in Social Mobility Tables". Behavioral Sciences, 7, pp. 14-27.

APPENDICE 1

LE PROGRAMME ANOMHI

```
)LOAD ANOMHI

)FNS


ANOMHI
FIN
PCOPY
PREP1A
PREP1B
PREP1C
REV
RND
TEMPS
▲CLEAR
```

LES PROGRAMMES RND, ▲CLEAR ET PCOPY SE TROUVENT EN LIBRAIRIE PUBLIQUE. LE PROGRAMME TEMPS EST UNE MODIFICATION DU PROGRAMME TIME INSCRIT EN LIBRAIRIE PUBLIQUE. LES PROGRAMMES REV ET PREP2 ONT ETE RESPECTIVEMENT ECRITS PAR M. EMILE BOUDREAU ET M. PAUL BELAND QUE NOUS REMERCIONS DE LEUR COLLABORATION.

```
    ∇ANOMHI[⎕]
  ∇ ANOMHI;LA;X;AX;N;C;X1;X2
[1]   0⍴⎕←'ANALYSE NOMINALE HIERARCHIQUE DE FREQUENCES'
[2]   'PROGRAMME ECRIT PAR:'
[3]   'FRANCOIS DELAND
      ESSION DU ';100 100 100⊤100 100 100⊥3↑⎕TS
[4]   'A.S.D.P.E.,1978                                DE
      BUTE A: ';TEMPS
[5]   0⍴⎕←'UNIVERSITE LAVAL'
[6]   0⍴⎕COPY'RESERVE PREP1 PREP1A PREP1B PREP1C'
[7]   PREP1
[8]   0⍴⎕EX 6 5⍴'PREP1 PREP1APREP1BPREP1C'
[9]   →('NON'≡3⍴⍞,⍞←'DESIREZ-VOUS LA LISTE DES TRAVAUX?')/18
[10]  'QUEL TRAVAIL DESIREZ-VOUS FAIRE EXECUTER?'
[11]  '1) UNE ANALYSE GENERALE D''UN TABLEAU DE CONTINGENCE
      MULTIVARIE?'
[12]  '2) UNE ANALYSE DE REGRESSION?'
[13]  '3) L''EVALUATION DES FREQUENCES ESTIMEES PAR UN OU DE
      S MODELES NON SATURES?'
[14]  '4) LA REPRODUCTION ET LA NORMALISATION DU (DES) TABLE
      AU(X) DE FREQUENCES EVALUE(S) PAR UN MODELE
[15]  '5) LE CALCUL DES COEFFICIENTS DE DETERMINATION MULTIP
      LE?'
[16]  '6) LE CALCUL DE(S) COEFFICIENT(S) DE DETERMINATION PA
      RTIELLE?'
[17]  '7) LE CALCUL DE(S) COEFFICIENT(S) D''EFFICACITE DE LA
       REDUCTION DE CATEGORIES?'
[18]  '(INSCRIVEZ LE NUMERO DE LA QUESTION QUI CORRESPOND AU
       TRAVAIL DESIRE)'
[19]  X2←0 0 0 0 0 0 0
[20]  →(A,B,C,D,E,F,G)[X1←⎕]
[21] A:0⍴⎕COPY'RESERVE PROG1 PREP2 PREP4 PREP5A PREP5B PREP6
      A PREP6B PREP7 PREP9 SOD1 SOD2'
[22]  →34
[23] B:0⍴⎕COPY'RESERVE PROG2 PREP2 PREP4 PREP5A PREP5C PREP6
      B SOD1 SOD2'
[24]  →34
[25] C:0⍴⎕COPY'RESERVE PROG3 PREP8A PREP8B PREP8C PREP8D PRE
      P8E PREP10 PREP11'
[26]  →34
[27] D:0⍴⎕COPY'RESERVE PROG4 PREP3A PREP3B PREP3C PREP3D PRE
      P3E PREP10'
[28]  →34
[29] E:0⍴⎕COPY'RESERVE PROG5'
[30]  →34
[31] F:0⍴⎕COPY'RESERVE PROG6'
[32]  →34
[33] G:0⍴⎕COPY'RESERVE PROG7'
[34]  X2[X1]←1
[35]  ⍎((4⍴1),X2)/'PROG1234567'
[36]  0⍴⎕EX 16 6⍴'PROG1 PROG2 PROG3 PROG4 PROG5 PROG6 PROG7
      PREP2 PREP3APREP3BPREP3CPREP3DPREP4 PREP5APREP5BPREP5C
      '
[37]  0⍴⎕EX 13 6⍴'PREP6APREP6BPREP7 PREP8APREP8BPREP8CPREP8D
      PREP8EPREP9 PREP10PREP11SOD1  SOD2  '
[38]  →('OUI'≡3⍴⍞,⍞←'DESIREZ-VOUS ENTRER D''AUTRES DONNEES?'
      )/6
[39]  →('OUI'≡3⍴⍞,⍞←'DESIREZ-VOUS FAIRE EXECUTER UN AUTRE TR
      AVAIL?')/18
[40]  FIN
  ∇
```

```
    ∇ FIN
[1]  'FIN DE LA SESSION A: '⍕TEMPS
[2]   'VOUS SORTEZ D''ANOMHI, VOUS ENTREZ DANS UN'
    ∇

    ∇ PCOPY

    ∇ PREP1A;C1
[1]  'LES FREQUENCES:'
[2]   'F8.2'⎕FMT((AX)⍴MD1)
[3]   →('NON'=3↑⍞,⍞←'VERIFIEZ LES FREQUENCES, AVEZ-VOUS COMM
      IS DES ERREURS?'')/0
[4]   'QUELLE(S) EST (SONT) LA (LES) VALEUR(S) DE LA (DES) F
      REQUENCE(S) A CORRIGER?'
[5]   C1←(MD1∊⎕)/⍳⍴MD
[6]   'PAR QUELLE(S) VALEUR(S) EST-ELLE (SONT-ELLES) REMPLAC
      EE(S)?'
[7]   MD1[C1]←⎕
[8]   ⍎((2⍴1),(C=0),4⍴1)/'MDA←MD1'
[9]   →1
    ∇

    ∇ PREP1B;QA;O;F;FA1;MDL;MOD;LMOD;X2;TM1;DM;TA;F1;I;J1;J1
      ;Q;DE;A;X5;FM
[1]   D←J1←⍴A←1-O←FM←1
[2]   O⍴PCOPY'RESERVE PREPBA PREPBB PREPBC'
[3]  '***ATTENTION: N''ENTREZ QU''UN SEUL MODELE***'
[4]   PREPBA
[5]   O⍴⎕EX 3 6⍴'PREPBAPREPBBPREPBC'
[6]   MDA←F
    ∇

    ∇ PREP1C;X1;X2;X3;X4;I;I1;I2;MOD;LMOD;LMD;MD1
[1]   X1←⍳⍴X2←X,⍴I←I1←I2←0
[2]   MOD←MD1←X⍴MD
[3]   LMOD←MDA←X⍴MDA
[4]   LMD←MDA
[5]   MD1←(X3←(1↑X1),(~(1↑X1)=⍳⍴X)/⍳⍴X)⍉(X4←X[1↑X1],X[(~(1↑X
      1)=⍳⍴X)/⍳⍴X])⍴+/[1↑X1+1⌽X1]MD1
[6]   MDA←X3⍉X4⍴+/[1↑X1]MDA
[7]   →(((⍴X)-1)≠I2←I2+1)/5
[8]   X1←1⌽X1,⍴I2←0
[9]   LMOD←LMOD×MD1÷MDA
[10]  X1←1⌽X1
[11]  X2←1⌽X2
[12]  MD1←MOD
[13]  MDA←LMOD
[14]  →((⍴X2)≠I←I+1)/5
[15]  I←0
[16]  →(~(∧/,((|LMD)-MDA)≤0.01)∨30≠I1←I1+1)/4
[17]  MDA←,MDA
    ∇
```

```
      ∇ MD←REV;X1
[1]     MD←⍳0
[2]     →4×⍳⍴X1←⎕,⍴0
[3]     →5
[4]     →2,0⍴MD←MD,X1
[5]     MD←MD
      ∇

      ∇ R←N RND X
[1]     R←(10*-N)×⌊0.5+X×10*N
      ∇

      ∇ C←TEMPS;T
[1]     T←2 12 60 60 60⊤⍳ 60 60 60⊥1↓(1 1 1,1000÷60)×3↑⎕TS
[2]     T←4⍴T
[3]     T[2]←T[2]+12×T[2]=0
[4]     T←100⊥T
[5]     T←10,(2,6⍴10)⊤T
[6]    C←('0123456789:'[1+T[3 4 1 5 6 1 7 8]]),' ',('AP'[1+T[2
        ]]),'M'
      ∇

      ∇ ⍝CLEAR
```

```
)LOAD RESERVE

)PNS

PREP1A
PREP1B
PREP1C
PREP10
PREP11
PREP1
PREP2
PREP3A
PREP3B
PREP3C
PREP3D
PREP4
PREP5A
PREP5B
PREP5C
PREP6A
PREP6B
PREP7
PREP8A
PREP8B
PREP8C
PREP8D
PREP8E
PREP9

PROG1
PROG2
PROG3
PROG4
PROG5
PROG6
PROG7
SOD1
SOD2
SOD3
```

```
    ∇ PREP10
[1]   II←' CYCLES ITERATIFS ONT ETE EXECUTES.'
[2]   'LE CRITERE D''AJUSTEMENT N''EST PAS ATTEINT...'
[3]   CL←1↑CL,IR←IR+B [illegible]
      ONTINUER LE PROCEDE ITERATIF'
    ∇

    ∇ PREP11
[1]   ' '
[2]   'ATTENTION, UNE OU DES CONFIGURATIONS CONTIENNENT'
[3]   'UN OU DES ZERO. CERTAINES CELLULES NE SONT PAS ESTIMEES.'
[4]   'CORRIGEZ LES DEGRES DE LIBERTE EN CONSEQUENCE.'
[5]   ' '
    ∇

    ∇ PREP1;X1;MD1;C;MDA;K;INT
[1]   →(1↑'NON'=3⍴⍞,⍞←'VOULEZ VOUS COPIER VOTRE VECTEUR DANS U
      N BLOC DE TRAVAIL? ')/4
[2]   'COPIEZ VOTRE VECTEUR:'
[3]   X1←⎕
[4]   'INSCRIVEZ VOTRE VECTEUR DE DONNEES:'⍴C←INT←1
[5]   MD1←MD←REV
[6]   'INSCRIVEZ LE NOMBRE DE CATEGORIES DE CHACUNE DE VOS VA
      RIABLES:'
[7]   AX←(X/ 1↓2), 1↑X←⎕
[8]   →(1↑,0 52↑((⍴⎕[illegible]×/A))/'ERREUR DE DIMENSION, R
      ECOMMENCEZ ET FAITES ATTENTION'),⍞0),0)/7
[9]   'INSCRIVEZ LES CARACTERES QUI REPRESENTENT VOS DONNEES
      '
[10]  LA←⍞
[11]  →(1↑,0 52↑((⍴(⎕←(((⍴X)≠⍴LA))/'ERREUR DE DIMENSION, RE
      COMMENCEZ ET FAITES ATTENTION'),⍞0),0))/9
[12]  PREP1A
[13]  →((C=0),0⍴⎕←[illegible]/'UNE OU DES FREQUENCES SON
      T PETITES, DESIREZ-VOUS UNE CORRECTION [illegible]
[14]  →('NON'=3⍴⍞)/28
[15]  ⍎(6⍴C←'OUI'=3⍴⍞,⍞←'LES PROBABILITES A PRIORI SONT-ELLE
      S DETERMINEES PAR UN MODELE LOG-LINEAIRE?')/[illegible]
[16]  →((C=1),0⍴(⎕←((⍳45)×((C←1↑C)=1))/'ENTREZ LES FREQUENCES
      D''UN AUTRE ECHANTILLON:')))/26
[17]  MD1←MDA←REV
[18]  PREP1A
[19]  PREP1C
[20]  →(INT=0)/31
[21]  MD1←(((+/MD)×MD)÷(K×MDA))÷(+/MD)+K×((+/MD)*2)-(+/MD*2)
      ÷+/(MDA-MD)*2
[22]  'LES FREQUENCES CORRIGEES:'
[23]  'F8.2'⎕FMT(AX⍴MD1)
[24]  ' '
[25]  'CONSTANTE K: ';K
[26]  ' '
[27]  ⍎(6⍴('NON'=3⍴⍞,⍞←'DESIREZ VOUS UTILISER LES DONNEES OR
      IGINALES?'))/'MD←MD1'
[28]  ' '
[29]  'FIN DE LA PROCEDURE D''ENTREE DU VECTEUR DE DONNEES:'
[30]  →0,⎕←⍴0
[31]  →('OUI'=3⍴⍞,⍞←'DESIREZ-VOUS ENTRER D''AUTRES DONNEES?'
      )/1
[32]  FIN
    ∇
```

```
    ∇ PREP2;X1;X2;X3;X4;X5;C
[1]   X2←X1←⍴X
[2]   B1←(((2×X1)-1),X1)⍴0
[3]   B1[⍳X1;1]←⍳X1
[4]   C←X3←0
[5]   →(((⍴B1)[1])=X3←X3+1)/0
[6]   C←0
[7]   →(X1=(X4[⍴X4←(B1[X3;]≠0)/B1[X3;]])/5
[8]   B1[X2←X2+1;]←(1,⍴X)⍴X4,(X1-⍴X4←X4,(X5←(X5[⍴X4]+C←C+1))
      )⍴0
[9]   →(X5<X1)/7
[10]  →(X3<(1 0/⍴B1))/5
    ∇

    ∇ PREP3A;X1;LA2;F1
[1]   →(MV2≠X1←1)/C
[2]   →((X1←∧/LA[C]=LA2←(LA1∊LA[C])/LA1)∨(1=⍴,(LX∊LA[C])))/0
      )/B
[3]   'ATTENTION, DEUX TABLEAUX SERONT PRESENTES POUR CHAQUE
       SORTIE.';⎕←⍴0
[4]   →((∧/LA=LA1)∧(1≠⍴LX∪LA))/A
[5]   'DANS LE PREMIER, L''ORDRE DES VARIABLES SERA ';LA[C];
      ' PLUTOT QUE ';LA2;'.'
[6]   'LA (LES) VARIABLE(S) DEPENDANTE(S) ';((LX∊LA2)/LX);' AP
      PARAITRA(ONT) EN COLONNES.'
[7]   'LE SECOND TABLEAU RESPECTERA LA FACON ORIGINALE D''EN
      TREE DES DONNEES.'
[8]  'LES FREQUENCES NORMALISEES:';⎕←⍴0
[9]  (16⍴~X1)/'PREMIER TABLEAU:'
[10]  'F10.2'⎕FMT(AX⍴F)
[11]  →X1/14,⎕←⍴0
[12] 'DEUXIEME TABLEAU:'
[13]  'F10.2'⎕FMT((AX1←(×/¯1↓AX1),¯1↑AX1)⍴(⍋LM[C])⍉X[C]⍴F)
[14] 'LES PROPORTIONS NORMALISEES:';⎕←⍴0
[15] (16⍴~X1)/'PREMIER TABLEAU:'
[16]  'F10.3'⎕FMT(AX⍴F1←⍎(1,1,1,(7⍴MV2),1,1,(3⍴MV2),1)/'F÷,⍉
      (⌽AX)⍴+/AX⍴F')
[17]  →X1/20,⎕←⍴0
[18] 'DEUXIEME TABLEAU:'
[19]  'F10.3'⎕FMT(AX1⍴(⍋LM[C])⍉X[C]⍴F1÷+/,F1)
[20]  →(MV1=0)/0,⎕←⍴0
[21] 'LES FREQUENCES OBSERVEES:'
[22] (16⍴~X1)/'PREMIER TABLEAU:'
[23]  'F10.2'⎕FMT AX⍴F2
[24]  →X1/0,⎕←⍴0
[25] 'DEUXIEME TABLEAU:'
[26]  'F10.2'⎕FMT(AX1⍴(⍋LM[C])⍉X[C]⍴F2)
    ∇
```

```
    ∇ PREP1A;C1
[1]   'LES FREQUENCES:'
[2]   'F8.2'⎕FMT((A×)⍴MD1)
[3]   →('NON'≡3↑B,⍞←'VERIFIEZ LES FREQUENCES. AVEZ-VOUS COMM
      IS DES ERREURS?')/0
[4]   'QUELLE(S) EST (SONT) LA (LES) VALEUR(S) DE LA (DES) F
      REQUENCE(S) A CORRIGER?'
[5]   C1←(MD1∊⎕)/⍳⍴MD
[6]   'PAR QUELLE(S) VALEUR(S) EST-ELLE (SONT-ELLES) REMPLAC
      EE(S)?'
[7]   MD1[C1]←⎕
[8]   ⍎((2⍴1),(C=0),4⍴1)/'MDA←MD1'
[9]   →1
    ∇

    ∇ PREP1B;DA;D;F;A;MIC;MOD;LMOD;X2;I1;I;DM;M;F1;1,J1;J1
      ;Q;DE;A;X5;FM
[1]   D←J1←⍴A←1-0←FM←1
[2]   0⍴⎕COPY'RESERVE PREFBA PREFBB PREFBC'
[3]  '***ATTENTION: N''ENTREZ QU''UN SEUL MODELE***'
[4]   PREFBA
[5]   0⍴⎕EX 3 6⍴'PREFBAPREFBBPREFBC'
[6]   MDA←F
    ∇

    ∇ PREP1C;X1;X2;X3;X4;I;I1;I2;MOD;LMOD;LMD;MD1
[1]   X1←⍳⍴X2←X,⍴I←I1←I2←0
[2]   MOD←MD1←X⍴MD
[3]   LMOD←MDA←X⍴MDA
[4]   LMD←MDA
[5]   MD1←(X3←(1↑X1),(~(1↑X1)=⍳⍴X)/⍳⍴X)⍉X4 X[1↑X1]+XI(1↑1+
      1)=⍳⍴X)/⍳X]⍴+/[1↑X1+1⌽X1]MD1
[6]   MDA←X3⍉X4⍴+/[1↑X1]MDA
[7]   →(((⍴X)-1)≠I2←I2+1)/5
[8]   X1←1⌽X1,⍴I2←0
[9]   LMOD←LMOD×MD1÷MDA
[10]  X1←1⌽X1
[11]  X2←1⌽X2
[12]  MD1←MOD
[13]  MDA←LMOD
[14]  →((⍴X2)≠I←I+1)/5
[15]  I←0
[16]  →(~(∧/,((|LMD)-MDA)≤0.01)∨30≠I1←I1+1)/4
[17]  MDA←,MDA
    ∇
```

```
    ∇ PREPARE[illegible]AX1;X1;DC1;LX;LM;F;I1;IR;F1;MV1;TM[illegible]C
[1]   DC[illegible]DC[illegible]X.OρLA1←LA.OρAX1←AX.OρCL←0
[2]   'QUEL[illegible] EST (SONT) LA (LES) VARIABLE(S) DEPENDANTE(
      S)[illegible]'
[3]   L[illegible] LA∊LX)/ιρX).(LA∊LX←□[illegible]/ιρX]
[4]   [illegible]3ρ□.□←'VOULEZ-VOUS NORMALISER LES INTERA
      CTIONS DE[illegible] VARIABLES INDEPENDANTES?')/15
[5]   [illegible]LA1[illegible])/X).*/(LA1[illegible])/X
[6]   F[illegible]LM⍉×ρDC)/.LM⍉XρMD
[7]   F[illegible]
[8]   [illegible]
[9]   F[illegible]DC/.AXρ+/[1]AXρ(F1←.F)
[10]  F[illegible].&(1⊖AX)ρ+/AXρ(.F)
[11]  →[illegible]I1[illegible]I1+1))/'PREP10'
[12]  [illegible]
[13]  [illegible](F÷+/F)×+/DC1/MD)⍴DC[illegible]/?
[14]  F[illegible]
[15]  →[illegible]ρ□.□←'VOULEZ-VOUS NORMALISER LES TABLEAUX D
      [illegible]TRES D''UN MODELE?')/'PREP30'
[16]  [illegible]AX1;OρLA←LA1;OρX←X1
```

```
    ∇ PREP3C;CL;X3;I;IR;MDA1;LX1;DLX;F;DC;X;A;MDA1;MDD;
      X2;X4;X7;X8;F1;DC2

[1]   F←DC\DC/MD,DLX←LX1←MDA1←ρIR←5↑1)←X3←CL←0
[2]   'COMBIEN D''ENSEMBLES DE MARGINALES DESIREZ VOUS ENTRE
      R?'⍞D←0ρDC1←XρDC

[3]   X6←⎕
[4]   'INSCRIVEZ UN ENSEMBLE DE MARGINALE(S):'
[5]   LX1←LX1,LX←⎕
[6]   'INSCRIVEZ LES VALEURS DES FREQUENCES DES MARGINALES:'
[7]   MDA1←MDA1,MDA←REV
[8]   X3←X3,ρMDA
[9]   →(X6≠ρ(DLX←DLX,+/LA∊LX),ρD)/4
[10]  →(I,0ρD←(12ρI←'OUI'=3ρ⍞,D←'AVEZ VOUS COMMIS DES ERREUR
      S?')/'RECOMMENCEZ:')/1

[11]  I←0
[12]  MDB←F
[13]  X4←X2←0,(~LA∊LX←(1↑DLX)↑LX1)/⍳ρX
[14]  DC1←XρDC
[15]  DC1←+/[1↑X4←1⌽X4]DC1
[16]  →(0=(X4←X4-1)[2])/15
[17]  MDA←DC\DC/,((X7/⍳ρX),X8/⍳ρX)⍉(((X7←~LA∊LX)/X),((X8←LA∊
      LX)/X))ρDC2\(DC2/(1↑X3←1⌽X3)↑MDA1)÷(DC2←~(,DC1)∊0)/,DC
      1
[18]  MDD←((1↑X2),X1)⍉(X[1↑X2],X[X1←(~(⍳ρX)∊1↑X2)/⍳ρX]))ρ+/[1
      ↑X2]Xρ(,MDB)÷X[1↑X2←1⌽X2]

[19]  →(0≠X2[2])/18
[20]  F←DC\(DC/(F1←F)×,MDA)÷DC/,MDB
[21]  MDA1←(1↑X3)⌽MDA1
[22]  LX1←(1↑DLX)⌽LX1
[23]  →((I←I+1)≠(ρDLX←1⌽DLX))/12
[24]  X3←1⌽X3
[25]  ⍎(6ρ(IR=I)+1)+1)))/'PREP1G'
[26]  →(CL=1)/28
[27]  →(∨/,(((1÷1-F)≥DL))/11
[28]  'LES FREQUENCES NORMALISEES:'
[29]  'F8.2'⎕FMT AXρF
[30]  ⍎(4ρCL←'OUI'=3ρ⍞,D←'VOULEZ VOUS CONTINUER VOS TRAVAUX
      AVEC LES FREQUENCES NORMALISEES?')/'MD←F'

[31]  (82ρ(CL=1))/'***VOTRE VECTEUR DE DONNEES EST MAINTENAN
      T CONSTITUE DES FREQUENCES NORMALISEES***'⍞D←ρD

    ∇
```

```
    ∇ PREP[illegible]X2;DM;C;TM2;T;TM1;DC;TM3;F2;X4;X3;[illegible]
      ;IR;[illegible];AX;AX;F1;X5
[1]   'INSCRIVEZ LE(S) PARAMETRE(S) D''INTERACTION DU MODELE:
      '[illegible]ρMV1+1
[2]   X2[illegible](F+X1∊',')))/X1+B
[3]   TM1+1+TM2+C+0,ρDM++/F+1,F
[4]   [illegible]TM2+TM2+ρρF+1⊖F
[5]   →([illegible])/4
[6]   →((C+1+TM2+0)≠(ρF+(ρρTM1+TM1,TM2[illegible])/4
[7]   [illegible](,DC)\(,DC)/MD),0ρDC1+DC+[illegible]DC
[8]   TM3+TM[illegible]+0,TM2+(((X1\(C+(X1+LA[illegible](ρLA)+((1↑T[illegible]
      )))/(ρLA))≠(ρLA)ρ0)/⍳ρLA)
[9]   →[illegible]/13
[10]  F2+F[illegible]/[1↑TM3+1⊖TM3]F
[11]  DC+[illegible]/[1↑TM3]DC)÷X[1↑TM2+1⊖TM2]))>0
[12]  →([illegible]TM3-1)[2])/10
[13]  '[illegible]'↑LA[C],D+'[illegible]
[14]  [illegible]X4+C
[15]  [illegible]I1+CL+0
[16]  I+1
[17]  [illegible]DC1)/MD)×F+∘/F+DC\(DC/,F)[illegible](DC+,F[illegible]((1↑X3)[illegible]
      [illegible]/C1)⍉(X[1↑X4],X[((1↑X4)≠C)/C])+∘/[1↑X3]×[illegible]
[18]  [illegible]
[19]  →((ρX3-1⊖X3)≠I+1+1)/17
[20]  ⍎([illegible]I1+I1+1))/'PREP10'
[21]  →([illegible])/24
[22]  I+0
[23]  →([illegible]((I[illegible],F1)[illegible],F)∧DE))/16
[24]  A[illegible]⍕((2ρMV2=0),3ρMV2-1)/'-1 X5')↓X[C]),×/(⍕((2ρMV2
      [illegible]ρMV2-1)/'-1 X5++/(LX∊LA[C],')↑X[C]
[25]  ⍎([illegible]MV2[illegible]1)/'AX1+(X[C])[⍋LM[C]]'
[26]  X2[illegible])⊖X2
[27]  DC+[illegible]
[28]  PREP3A
[29]  →([illegible]+1)/7
    ∇

    ∇ PREP[illegible]X1;X2;X3
[1]   X3+,A[illegible](X1+1↑X+¯1⊖X
[2]   C+ρρMD
[3]   X3+X3,,((AX)ρ⍉((×/X1),(ρX2))ρX2+,1↑X+¯1⊖X)
[4]   →((ρX)≠C+L+ρρX1+X1,ρX2)/3
[5]   J+(⍉((ρX),(×/AX))ρX3)[⍋⍳ρX]
    ∇
```

```
    ∇ PREP5A;X1;X2;X3;X4;X5;L;LB;[illegible];T1;MDL;T2;C;C1;TA;MDL1;[illegible]
      PS;PS1;PS2;MDL2
[1]   C←C1←LB,⍬LB←(T1←1),ρX2←MDL1←PS1←ρMDL←0
[2]   X5←ρX1←0,X[M←(,B1[COM;]≠0)/,B1[COM;]]
[3]   X2←X2,X3←(1↑X1←1⊖X1)
[4]   →((1↑1⊖X1)≠0)/3
[5]   X1←1⊖X1
[6]   X3←(1↑X1←1⊖X1)↑X2
[7]   X4←X3←(10×X3)∘.+X4←(1↑[illegible]1⊖X1)↑X2←(1↑X1)⊖X2
[8]   →((¯1↑2↑X1)≠0)/7
[9]   ⍎((6ρ((OPUS=1)∨((CE=1)∧((¯1↑M)=ρX1)[illegible]
      CE=1)∧((¯1↑M)=ρX)))/'PREP6MEKLEA[illegible]'
[10]  LB←(LB×10)∘.+L←1 0
[11]  →((ρX[M])≠(T1←T1+1))/10
[12]  T1←0
[13]  T2←ρρPS←(ρMD)ρ0
[14]  T←T-ρT←(P←,∧/(T/J[;M])=(ρT/J[;M])ρ((T←,(((ρX[M])ρ10)⊤[illegible]
      ↑,LB))/,(((ρX[M])ρ10)⊤TA←1↑,X4)))/[illegible]
[15]  →((ρC)≠T2-1)/17
[16]  C1←((ρC)-2)↓C←C,⍬C
[17]  MDL←((1↑C1),(¯1↑C1))/(MDL+[illegible]/T),(MDL[illegible]
[18]  PS←(((ρMD)ρ(1↑C1)),((ρMD)ρ(¯1↑C1)))/(PS+P×(ρMD[illegible]
      S-P×(ρMD)-ρT)
[19]  →((T2←T2+ρρLB←1⊖,LB)≠(ρ,LB×ρρC1←1↓C1))/14
[20]  MDL←((1↑C1),(¯1↑C1))/(MDL++/((MP≠0)/MP)÷ρMD),(MDL-+/((
      MP≠0)/MP)÷ρMD)
[21]  PS←((((ρMD)ρ(1↑C1)),((ρMD)ρ(¯1↑C1)))/(PS+(PS2×P1)),((
      (ρMD)ρ1))÷ρMD
[22]  →(OPUS=0)/28
[23]  ⍎((6ρ((10↑TA)=1)),(5ρ((10↑TA)≠1)))/'PS2←PSPS←PS'
[24]  ⍎((8ρ((10↑TA)=1)),(7ρ((10↑TA)≠1)))/'MDL2←MDLMDL←MDL'
[25]  →((10↑TA)=1)/32
[26]  MDL←(MDL2-MDL)÷2
[27]  PS←(PS-PS2)÷2
[28]  MDL1←MDL1,MDL
[29]  PS1←PS1,PS
[30]  P←(+/((MP≠0)/PS*2)÷(MP≠0)/MD+0.5)*0.5
[31]  ⍎((5ρ((+/LQ∊LA[M])=1)),(4ρ((+/LQ∊LA[M])≠1)))/'PREP5DD
      2'
[32]  C1←C←C,⍬C←1,MDL←0
[33]  ⍎(11ρ((+/LQ∊LA[M])≠1))/'PS1←MDL1←ρ0'
[34]  →((T1←T1+ρρLB←(ρρX4←1⊖,X4)⊖,LB)≠(ρX4))/13
    ∇
```

```
    ∇ [illegible]DL;X2;TA;MDL1;PS1;P;PS
[1]   MDL1←PS1←⍳0
[2]   X1←[illegible],X1←((X[M])-(("∨/LQ∊LA[M]),∨/LQ∊LA[M[illegible](,B1[COM;]≠0)
      /,B1[COM;]])/1 0
[3]   →(((C[illegible]1)∧(M=⍴X))/0
[4]   MDL←[illegible]DL-⍴MDL+(PS←,(J[;M])=(⍴J[;M])⍴X2←TA←¯1↑X1←1⊖X1)
      /MP)-(+,((MP≠0)/MP)-⍴MD)
[5]   MDL1←[illegible]1,MDL
[6]   P[illegible]1,PS←((PS×(⍴MD)÷⍴PS/PS)-(⍴MD)⍴1)-⍴MD
[7]   P←[illegible]MP≠0)/PS*2)÷(MP≠0)/MD+0.5))*0.5
[8]   →([illegible]LA[M])∧((1↑1⊖X1)≠0))/4
[9]   ⍎([illegible]Q∊LA[M]),4⍴("∨/LQ∊LA[M]))/'PREP2SOD2'
[10]  →([illegible]1↑(⊖×1)≠0)/4
    ∇

    ∇ [illegible]X1;MDL;X2;TA;MDL1;P;P1;PS
[1]   X[illegible]((X[M;(,B1[COM;]≠0)/,B1[COM;]])
[2]   [illegible]L-⍴MDL+(PS←,(J[;M])=(⍴J[;M])⍴X2←TA←1↑X1←1⊖X1)
      [illegible]((MP≠0)/MP)-⍴MD)
[3]   →([illegible]1)≠0)/7
[4]   [illegible](⍴MD)÷⍴PS/PS)-(⍴MD)⍴1)
[5]   [illegible]
[6]   →[illegible]
[7]   [illegible](MP≠0)/(P1-((PS×(⍴MD)÷⍴PS/PS)-(⍴MD)⍴1))÷2×⍴MD)
      +[illegible])/MD+0.5))*0.5
[8]   [illegible]-MDL)÷2
[9]   SOD[illegible]
[10]  →([illegible]1)≠0)/2
    ∇

    ∇ [illegible]
[1]   X[illegible](X[M])[X5]-1)⍴1),0)/[X5←X5-1]X4
[2]   →([illegible]X4)≠(×/(X[M]-1)))/1
    ∇

    ∇ PREP61
[1]   X5←X5-2
[2]   →⍎((2∨/OPUS=1),OPUS=0)/'((CE=1)∧(((⍴X)-+/CX∊LA)∊M))/43
      '
[3]   X4←,((((X[M])[X5]-1)⍴1),0)/[X5]X4
[4]   X5←X5-1
[5]   →(X5≠0)/3
    ∇

    ∇ PREP2
[1]   MDL←(1↑MDL1-¯1↑MDL1)÷2
[2]   P←(+/(((MP≠0)/(((⍴MD)⍴PS1)--(⍴MD)⍴PS1)÷2)*2)÷(MP≠0)/MD
      +0.5)*0.5
[3]   IN←1
[4]   SOD2
[5]   →((⍴MDL1)=2)/0
[6]   MDL←-/(IN×MDL1)÷+/IN←(0,⍳(⍴MDL1)-1)∘.((⍴MDL1)-1)
[7]   P←(+/((-/(((MP≠0)/((⍉((⍴PS),⍴IN)⍴IN)×((⍴IN),⍴PS)⍴PS1))
      ÷+/IN))*2)÷(MP≠0)/MD+0.5)*0.5
[8]   IN←2
[9]   SOD2
    ∇
```

```
    ∇ PREPBA;T;C;TM2;C1
[1]   MDL←MOD←TM1←X2←TM←T←⍴DM←C←TM2←C1←1 LMOD←F1←D 1
[2]   →(D=0)/6
[3]   →(F1←D←1⍴'NON'=(3⍴⍞,D←'VOULEZ-VOUS VERIFIER UNE (DES)
      HYPOTHESE(S) DE QUASI INDEPENDANCE?'))/6
[4]  'ENTREZ LE(S) VECTEUR(S) DE 1 ET DE 0:'
[5]   F1←REV
[6]  'SPECIFIEZ LE(S) CRITERE(S) D''AJUSTEMENT ITERATIF:'
[7]   J1←0,J1←⎕
[8]   →(FM=1)/10
[9]   X5←1⍴'OUI'=3⍴⍞,D←'VOULEZ-VOUS L''IMPRESSION DES ESTIME
      S DES FREQUENCES?'
[10] 'INSCRIVEZ LA (LES) MARGINALE(S) DU (DES) MODELE(S) QUE
      VOUS VOULEZ TRAITER:'
[11]  '(INSCRIVEZ UN SEUL PROBLEME PAR LIGNE)'
[12]  PREPBB
    ∇

    ∇ PREPBB;X1;X3;X4;K
[1]   X3←(~(F←X1,' ',' '))/X1←⍞
[2]   X2←X2,X3
[3]   MOD←MOD,X1
[4]   LMOD←LMOD,X4←⍴X1,⍴' '
[5]   DM←DM,+/F←1,F
[6]   C←C+⍴⍴TM2←TM2←⍴PF←1⍴F
[7]   →((F[2])≠1)/6
[8]   →((C+1+TM2←0)≠⍴PF←(⍴,TM1)⍴TM1,TM2)⍴F (+/F) )/6
[9]   TM←TM,TM1
[10]  →(((⍴J1)-1,TM1←⍴C←0)≠C1←C1+1)/1
[11]  ⍎(K←6⍴'NON'=(3⍴⍞,D←'VERIFIEZ VOS ENTREES. AVEZ VOUS CO
      MMIS DES ERREURS?'))/'PREPBC'
[12]  →(K=1)/0
[13]  'RECOMMENCEZ ET CETTE FOIS FAITES ATTENTION'
[14]  PREPBA
    ∇
```

```
    ∇ PREPBC;IR;CL;M;JI;B1;MOD1;MV;JE;I1;X1;DE;DE1;DE2
[1]   DE1←DE←F←⍎((6⍴1),(2⍴Q=0),(Q=1))/'(⍴MD)⍴F11'
[2]   J1←1⊖J1
[3]   IR←5←I←I1←C←CL←0
[4]   X1←(+/TM1←(1↑DM←1⊖DM,⍴TM)↑X2
[5]   TM2←0,TM2←(((MV\(MV←LA∊((⍴LA)⍴((1↑TM1)↑X1))))/⍳⍴LA)≠(⍴L
      A)⍴0)/⍳⍴LA)

[6]   →⍎((7⍴(⍴TM2)=1),(⍴TM2)≠1)/'21,F←MD?'
[7]   ⍎((9⍴1),(C=2),2⍴(C←C+1)=1)/'JI←DE\DE/TMD'
[8]   JI←,((1↑TM2),(((1↑TM2)≠⍳⍴X)/⍳⍴X))⍉(X[1↑TM2]∘.X[(1↓TM2)
      ≠⍳⍴X)/⍳⍴X]))⍴+/[1↑TM2←1⊖TM2]X⍴,JI

[9]   →(TM2[2]≠0)/8
[10]  TM2←1⊖TM2
[11]  J1←J1,JI
[12]  →(C=1)/7
[13]  DE1←DE\DE/DE2\(DE2←DE\~0=DE/(⍴MD)⍴J1)/DE1
[14]  C←1-⍴⍴F←DE1\(DE1/(JE←F)×(⍴MD)⍴J1)÷DE1/(⍴MD)⍴J1
[15]  X1←(1↑TM1)⊖X1,J1←⍴0
[16]  →((1↑DM)≠I←I+⍴⍴TM1←1⊖TM1)/5
[17]  ⍎(6⍴(IR=I1←I1+1))/'PREF10'
[18]  →(CL=1)/21
[19]  I←0
[20]  →(∨/,((|F-JE)≥1↑J1))/5
[21]  →(((D=0)∧(A=2))∨(FM=1))/0
[22]  'MODELE VERIFIE: ';(1↑LMOD←1⊖LMOD)↑MOD
[23]  ⍎(6⍴∨/0∊DE/F)/'PREF11'
[24]  →(X5=0)/27
[25]  'FREQUENCES ESTIMEES:'
[26]  'F8.2'⎕FMT((AX)⍴F)
[27]  'KHI DEUX USUEL: ';2 RND(+/((DE1/MD-F)*2)÷DE1/F)
[28]  'KHI DEUX DU RAPPORT DE VRAISEMBLANCE: ';2 RND(2×(+/(D
      E1/MD)×((~(DE1/MD)∊0)\⍟((~(DE1/MD)∊0)/(DE1/MD)÷DE1/F))
      )))
[29]  ' '
[30]  →(X5=0)/33
[31]  'DEVIATIONS DES ESTIMES:'
[32]  'M8-BF8.2'⎕FMT((AX)⍴DE1\DE1/MD-F)
[33]  ' '
[34]  →(D=0)/0
[35]  MOD←(1↑LMOD)⊖MOD
[36]  X2←(+/TM1)⊖X2
[37]  TM←(1↑DM)⊖TM
[38]  F1←(⍴MD)⊖F1
[3    →(J1[2]≠0)/1
    ∇
```

```
∇ PREPBD;MDA;MDL1;MDL2;LX;LMODA;J1A;MODA;X2A;TMA;DMA;[illegible]
  BR;RA;MDL3;MDL4;DE1;DE2;CH;R;C

[1]   CH←X3←RE←RA←1-LX←1
[2]   '1) HYPOTHESE(S) D''INDEPENDANCE GEN[illegible]'
[3]   PREPBA
[4]   ⍎(6ρI≠0)/'PREPBC'
[5]   FA←FA,F
[6]   ⍎(11ρ(A=2)∨(LX=0))/'LMOD←1⊖LMOD'
[7]   MOD←(1↑LMOD)⊖MOD
[8]   X2←(+/TM1)⊖[illegible]
[9]   TM←(1↑DM)⊖T
[10]  →(LX←∧/(4ρ((1↑LMOD)↑MOD))=[illegible])/[illegible]
[11]  TM1←(1↑DM←1⊖DM)ρTM
[12]  →ρ5ρJ1←1⊖J1
[13]  →I←(J1[2]≠0)/4
[14]  →((RA=1),RA=2)/26 28
[15]  0ρLMODA←LMOD←1⊖LMOD;0ρJ1A←J1←1⊖J1;0ρDMA←DM←1⊖[illegible]MODA
      ←MOD;0ρX2A←X2;0ρTMA←TM;FA←0,MDL2←FA

[16]  '2) HYPOTHESE(S) D''INDEPENDANCE REGIONALE(S) [illegible]'
[17]  →(1(~RE)-ρρ(⍎((5ρ1),2ρRE=1)/'MD←MDL4 [illegible]'
[18]  'COMBIEN D''HYPOTHESES D''INDEPENDANCE REGIONALE VOULE
      Z-VOUS VERIFIEZ?'

[19]  CH←0,ιCH←⎕
[20]  'HYPOTHESE REGIONALE NO. ';1↑CH←1⊖CH,⎕←0ρMDA←MD
[21]  PREPBE
[22]  →('NON'≡(3ρ⍞,⍞←'VERIFIEZ VOS ENTREES. AVEZ-VOUS COMMIS
       DES ERREURS?'))/24

[23]  →21,0ρ⍞←'RECOMMENCEZ ET CETTE FOIS FAITES ATTENTION',0
      ρMD←MDA

[24]  0ρLMOD←LMODA;0ρJ1←J1A;0ρDM←DMA;0ρMOD←MODA;0ρX2←X2A;0ρT
      M←TMA

[25]  →2+A←I←2
[26]  MDL3←(~0=FA)\((~0=FA)\MDL2)÷(~0=FA)/FA
[27]  →1+A←RA←ρ0,I←0,0ρ⍞←'STRUCTURE(S) D''HYPOTHESE DU TABLE
      AU REGIONALISE',⎕←FA←ρ0

[28]  R←ρρρMDL1←(((ρFA)↑ρMD),(ρMD))ρFA×,(ρFA)ρMDL3
[29]  LMOD←1⊖LMOD
[30]  'STRUCTURE D''HYPOTHESE DU TABLEAU REGIONALISE: ';(1↑LM
      OD←1⊖LMOD)↑MOD
```

```
[31]  ⍎([illegible]/DE1←0=MDL1[R;⍳⍴MD])/'PREP11'
[32]  →([illegible])/35
[33]  MDA←MD 2
[34]  '[illegible]'⎕FMT(AX⍴MDL1[R;⍳⍴MD])
[35]  'KHI DEUX USUEL: '⍕2 RND(+/(DE1/((MDA-,MDL1[R;⍳⍴MD])*2)
      )[illegible]DE1[illegible],DE1)/,MDL1[R;⍳⍴MD])

[36]  'KHI DEUX DU RAPPORT DE VRAISEMBLANCE: '⍕2 RND(2×(+/MDA
      [illegible]⍟(DE2/MDA)÷(DE2\DE1∧~MDA=0)/,MDL1[R;⍳⍴MD]))))

[37]  ' '
[38]  →([illegible])/41
[39]  'DEVIATIONS DES ESTIMEES:'
[40]  '[illegible]2'⎕FMT(AX⍴DE1\DE1/MDA-,MDL1[R;⍳⍴MD])
[41]  MOD[illegible]MOD)⍉MOD
[42]  ' '
[42]  [illegible]1↑[illegible]/MDL1)≠R←R+1)/30
[43]  →[illegible]'OUI'=3⍴⎕,⍞←'VOULEZ VOUS VERIFIER D''AUTRES MO
      DELES AVEC LA MEME REGIONALISATION?')/47,FA←⍴0

[45]  ⍎([illegible]A←~RA←BR←1⍴'OUI'=3⍴⎕,⍞←'VOULEZ-VOUS VERIFIER LE
      [illegible] MODELES D''INDEPENDANCE GENERALISEE?')/'MD←MDA
      [illegible]4←MD'
[46]  →([illegible],BR)/2 27
[47]  MD←MDA
[48]  →([illegible])/0
[49]  →([illegible]RA),RA←1⍴'OUI'=3⍴⎕,⍞←'VOULEZ-VOUS VERIFIER LES MEM
      ES MODELES D''INDEPENDANCE GENERALISEE?')/2 20

    ∇

    ∇ PREPBE;N;B1;B2;B3;MD3;MD4;R4;MD2;X1;MD5;C;X2;X3
[1]   MD2←X⍴MD,⍴C←0
[2]   'QUELLE EST LA VARIABLE TRAITEE?'
[3]   →((1↑⍴LA)=N←LA⍳⎕)/20
[4]   B1←((X1←⍳⍴X)[N])-1)⍴';'
[5]   'INSCRIVEZ LES GROUPES DE CATEGORIES:'
[6]   X3←X2←'0',⍞
[7]   C←C,('0'=1↑X2)∨','=1↑X2←1⌽X2
[8]   →(('0'≠1↑X2)∧','≠1↑X2)/7
[9]   B2←C/1⌽X3
[10]  B3←B2,((⍴LA)-X1[N])⍴';'
[11]  MD5←MD3←(+/[X1[N]]MD4←MD3←⍎'MD2[',B1,B3,']')÷⍴⍴B2
[12]  R4←⍳⍴MD4
[13]  R4[X1[N]]←1
[14]  MD3←R4⍴MD3
[15]  MD3←MD3,[X1[N]]R4⍴R4⍴MD5
[16]  →((×/⍴MD4)≠×/⍴MD3)/15
[17]  MD3←(⍴MD4)⍴MD3
[18]  ⍎'MD2[',B1,B3,']←MD3'
[19]  →(('0'=1↑X2),'0'≠1↑X3←X2)/2 7,⍴C←0
[20]  MD←,MD2
    ∇
```

```
    ∇ PREP9
[1]   →((ρMDL1,ρO)≠(((LA[M]≥LG)/ρ[M]) 1,1)/0
[2]   MDL1←MDL1,+/ MDL1
[3]   PREP7
[4]   PS1←MDL1÷ρO
    ∇

    ∇ PROG1;COM;T1;JE;I;LA1;JI;IN;F;I1;A;OPUS;A1;K;A;E;EI;A
      D2;AX1;XA1;JA1;EC;XB;LG;IN;JAC
[1]   PREP4
[2]   Oρ⎕←'***ANALYSE GENERALE D''UN TABLEAU DE FREQUENCES**
      *'
[3]   CX←LA[ρ(ρX)ρCE←EC←1;('NON' 3,0,⍞ 'TRAITEZ VOUS UN SEUL
       ECHANTILLON?')]
[4]   OPUS←+/2ρ('OUI' 3ρ⍞,⍞ 'LES CATEGORIES D''UNE VARIABLE
      SONT-ELLES EQUIDISTANTES?')
[5]   LG←'11'
[6]   →(OPUS≠2)/9
[7]   'QUELLE EST CETTE VARIANCE?'
[8]   LG←⎕
[9]   SOD1
[10]  PREP2
[11]  IN←OPUS←0,ρCOM←1
[12]  MP←(MD≠0)\⍟((MD≠0)/MD)-+/MD
[13]  ⍎((((ΔρCOM) (ρ((ρX)+1))),(2ρCOM)ρ(6ρ(ρX)+1)),'PROGS
      BPREPSA')
[14]  →(((( 1↑X)≠1))∧((2+(ρX))≠COM←COM 1 ))/17
[15]  ⍎(29ρ(CE=1))/'MD2←MD,OρXA1←( 1↑ρB),2,JA1←J',...
[16]  →(EC=0)/27
[17]  →(((ρXA1)+2)=CE←CE+1)/26
[18]  'ECHANTILLON NO. ',⍕CE-1
[19]  ⍎(18ρ((CE-1)=1))/'X← 1↓X1←X,OρAX1←AX',OρD←16
[20]  ⍎(5ρ((ρX)=1))/'X←X,1'
[21]  →((CE-1)≠1)/24
[22]  AX←(×/ 1↓X), 1↑X
[23]  PREP4
[24]  MD←,((AX1)ρMD2)[;1↑XA1]
[25]  →9,OρXA1←1↓XA1
[26]  MD←MD2,OρAX←AX1,OρX←XB,OρJ←JA1
[27]  ' '
    ∇
```

```
    ∇ PROG2;COM;T1;JE;P;LA1;J1;DM;MP;B1;M.[illegible];D1A;MD1;LA2;[illegible]
      X;LM;JA;X1;X2;X3;CE;MD2;AX1;X[illegible];XA[illegible];EC;[illegible];LO
[1]     PREP4
[2]     0⍴⎕←'***ANALYSE DE REGRESSION***'
[3]     'QUELLE(S) EST (SONT) LA (LES) VARIABLE(S) DEPENDANTE(
        S)?'
[4]     LM←((~LA∊CX)/⍳⍴X),(LA∊CX←⍞)/⍳⍴X
[5]     CE←EC←1⍴('NON'≡3⍴⍞,⎕←'TRAITEZ VOUS UN SEUL ECHANTILLON
        ?')
[6]     LO←'11',0⍴IN←0
[7]     0⍴XA1←X,0⍴LA2←LA,0⍴JA←J
[8]     X←X[LM],0⍴LA←LA[LM]
[9]     SOD1
[10]    PREP2
[11]    B1←B1[D1A/⍳⍴[illegible]A←∨/B1∘.[illegible](LA∊CX)/⍳⍴X)];]
[12]    OPUS←1↑1,COM←1,0⍴J←J[;LM]
[13]    MP←(MD≠0)\⌹((MD≠0)/MD)∘.+/MD
[14]    ⍎((6⍴(COM=+/LA∊CX)),6⍴(COM≠+/LA∊CX))/'PREP5PREP[illegible]'
[15]    →(((1↑⍴B1)+1)≠COM←COM+1)/14
[16]    ⍎(29⍴(CE=1))/'XA2←⍳¯1↑XA1,0⍴AX1←AX,0⍴MD2←MD',[illegible]
[17]    0⍴LA←LA2,0⍴X←XA1,0⍴J←JA
[18]    →(EC=0)/27
[19]    →(((⍴XA2)+2)=CE←CE+1)/26
[20]    'ECHANTILLON NO. ';CE-1
[21]    0⍴AX←(×/¯1↓X),¯1↑X←¯1↓X,0⍴LA←¯1↓LA
[22]    PREP4
[23]    MD←,((AX1)⍴MD2)[;1↑XA2]
[24]    LM←((~LA∊CX)/⍳⍴X),(LA∊CX)/⍳⍴X
[25]    →8,0⍴XA2←1⌽XA2
[26]    0⍴AX←AX1,0⍴MD←MD2
[27]    ' '
    ∇

    ∇ PROG3;O;F;FA;FA1;MDL;MOD;LMOD;X2;TM1;DM;TM;F1;I;J1;J2;
      Q;DE;A;X5;FM

[1]     J1←FA←⍴A←FM←0
[2]     0⍴⎕←'***EVALUATION DES FREQUENCES ESTIMEES PAR UN OU D
        ES MODELES NON-SATURES***'

[3]     ⍎((6⍴~O),6⍴O←1⍴'NON'=(3⍴⍞,⎕←'VOULEZ-VOUS VERIFIER UNE
        (DES) HYPOTHESE(S) D''INDEPENDANCE REGIONALE?'))/'PREP
        8DPREP8A'
    ∇
```

—

```
    ∇ PROG4;MV1;DC;DE;LX;MV2;LA1;X1
[1]   '***PROCEDURE ITERATIVE DE NORMALISATION DES FREQUENCE
      S***';D←0⍴MV1←0;0⍴LX←1↑LA1←LA
[2]   →(1↑DC←(⍴MD)⍴'NON'≡3⍴⍞,⍞←'VOTRE TABLEAU COMPTE-T-IL DE
      S ZERO STRUCTURAUX?')/5,D←⍴0
[3]   'ENTREZ LE VECTEUR DE 1 ET DE 0:'
[4]   DC←REV
[5]   'SPECIFIEZ LE CRITERE D''AJUSTEMENT ITERATIF:'
[6]   DE←⎕
[7]   →((0=∧/DC)∨(0∊MD))/11
[8]   ⍎(6⍴MV2←'OUI'≡3⍴⍞,⍞←'VOULEZ-VOUS AJUSTER VOS FREQUENCE
      S A CELLES D''UN AUTRE ECHANTILLON?')/'PREP3C'
[9]   →(MV2=0)/11
[10]  →('NON'≡3⍴⍞,⍞←'VOULEZ-VOUS NORMALISER LES TABLEAUX DE
       FREQUENCES EN INTERACTION?')/0
[11]  ⍎((~X1),X1←6⍴MV2←1⍴'OUI'≡3⍴⍞,⍞←'TRAITEZ-VOUS UN MODELE
       DE REGRESSION?')/'PREP3DPREP3D'
    ∇

    ∇ PROG5;X1;X2;X3
[1]   0⍴⍞←'***COEFFICIENTS DE DETERMINATION MULTIPLE***'
[2]   'ENTREZ VOS DONNEES:'
[3]   'A) HYPOTHESE D''EQUIPROBABILITE:'
[4]   'KHI DEUX:    DEGRES DE LIBERTE:'
[5]   X1←⎕
[6]   'B) HYPOTHESE D''INDEPENDANCE:'
[7]   'KHI DEUX:    DEGRES DE LIBERTE:'
[8]   X2←⎕
[9]   'C) HYPOTHESE A VERIFIER:'
[10]  'KHI DEUX:    DEGRES DE LIBERTE:'
[11]  X3←⎕
[12]  'COEFFICIENT DE DETERMINATION MULTIPLE: ';4 RND(((X1[1]
      ÷X1[2])-(X3[1]÷X3[2]))÷(X1[1]÷X1[2]))
[13]  'COEFFICIENT DE DETERMINATION MULTI-PARTIELLE: ';4 RND
      ((X2[1]÷X2[2])-(X3[1]÷X3[2]))÷(X2[1]÷X2[2]))
    ∇
```

```
    ∇ PROG6;C;X1;X2;X3;X4
[1]   0ρ⍞←'***COEFFICIENT(S) DE DETERMINATION PARTIELLE***'
[2]   X3←X4←ρ0
[3]   C←1↑('OUI'∊(3ρ⍞,⍞←'VOULEZ-VOUS VERIFIER LA CONTRIBUTIO
      N D''UN (DE) PARAMETRE(S) INCLUS?'))

[4]  'HYPOTHESE CONTROLE:'
[5]  'KHI DEUX:    DEGRES DE LIBERTE:'
[6]   →((C=1),C=0)/14 10
[7]  'HYPOTHESE(S) A VERIFIER:'
[8]  'KHI DEUX:    DEGRES DE LIBERTE:'
[9]   →(C=0)/14
[10]  X1←⎕
[11]  →(X1=0)/18
[12]  X4←X4,X1
[13]  →((C=0),(C=1))/7 9
[14]  X2←⎕
[15]  →(X2=0)/18
[16]  X3←X3,X2
[17]  →((C=0),(C=1))/9 7
[18] 'NO. DU PROBLEME:       COEFFICIENT(S) DE'
[19] '                    DETERMINATION PARTIELLE:'
[20]  'F7.0,M0-0F26.4'⎕FMT((⍳((1↑ρX4,ρ0)×(1↑ρX3,ρ0))),((X4 ∘
      X3←+/X3←((((ρX3)÷2),2)ρX3)))÷(X4←+/X4←((((ρX4)÷2),2)ρX
      4))))
    ∇

    ∇ PROG7;X1;X2;X3
[1]   D←X2←0ρ⍞←'***COEFFICIENT(S) D''EFFICACITE D''UNE REDUC
      TION DE CATEGORIES***'
[2]  'HYPOTHESE D''INDEPENDANCE GENERALISEE:'
[3]  'KHI DEUX:    DEGRES DE LIBERTE:'
[4]   X1←⎕
[5]  'HYPOTHESE(S) D''INDEPENDANCE REGIONALE:'
[6]  'KHI DEUX:    DEGRES DE LIBERTE:'
[7]   X2←X2,⎕
[8]   →((¯2↑X2)≠0 0)/7
[9]   X1←(((ρX2←¯2↓X2)÷2),2)ρX1
[10]  X2←(((ρX2)÷2),2)ρX2
[11] 'NO. DU PROBLEME:       COEFFICIENT(S) D''EFFICACITE:'
[12]  'F7.0,M0-0F26.4'⎕FMT(⍳1↑ρX2;((X1[;1]-X2[;1])÷X1[;1])×X
      2[;2]÷X1[;2]-(2×ρX)-(ρX)+1)
    ∇
```

```
    ∇ MD←REV;X1
[1]   MD←⍳0
[2]   →4×⍳⍴X1←⎕
[3]   →5
[4]   →2,0⍴MD←MD,X1
[5]   MD←MD
    ∇

    ∇ R←N RND X
[1]   R←(10*-N)×⌊0.5+X×10*N
    ∇

    ∇ SOD1
[1]   ' '                                                    TAU          LAM
[2]   ' PARAMETRES
     BDA          LAMBDA'
[3]   'RELATIONNELS
              STANDARDISES'
    ∇

    ∇ SOD2;N;LQ1
[1]   LQ1←((LQ∊LA[M])/LQ)
[2]   →((+/LQ∊LA[M])=1)/4
[3]   LQ1←(2×⍴LA[M],⍴' ')⍴'.'
[4]   N←((2⍴(IN=1)),2⍴(IN=2))/' L Q'
[5]   →(IN=2)/7
[6]   TA←(-(⍴X[M]))↑,TA←'I15'⎕FMT TA
[7]   '30A1,3M0-0F13.4'⎕FMT(1 30⍴LA1,N,(30-⍴LA1←(~''1⍉LQ1=LA1
      )/LA1←,⍉(2,(⍴TA,⍴0))⍴LA[M],TA)⍴'.',÷2.71828*MD[;MD;MD
      -P)
[8]   N←⍴IN←0
    ∇

    ∇ SOD3;N;LQ1
[1]   LQ1←((LQ∊LA[M])/LQ)
[2]   →((+/LQ∊LA[M])=1)/4
[3]   LQ1←(2×⍴LA[M],⍴' ')⍴'.'
[4]   N←((2⍴(IN=1)),2⍴(IN=2))/' L Q'
[5]   →(IN=2)/7
[6]   TA←(-(⍴X[M]))↑,TA←'I15'⎕FMT TA
[7]   '30A1,3M0-0F13.4'⎕FMT(1 30⍴LA1,N,(30-⍴LA1←(~''1⍉LQ1=LA1
      )/LA1←,⍉(2,(⍴TA,⍴0))⍴LA[M],TA)⍴'.',÷2.71828*MD[;MD;MD
      -P)
[8]   N←⍴IN←0
    ∇
```